Actin: Computational Design of Novel Inhibitors and Study of the Influence of the Natural Product Latrunculin on Conformational Dynamics

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Sandra Rennebaum
aus
Deutschland

Promotionskomitee

Prof. Dr. Amedeo Caflisch (Vorsitz)
Prof. Dr. Cristina Nevado

Zürich 2012

Summary

In the present thesis computational methods are employed to study actin, a protein constituting a major component of the eukaryotic cytoskeleton. Actin is an interesting drug target for the treatment of cancer since organization and polymerization state of the actin network are malignantly modified in tumor cells. Moreover, a local disruption of the actin cytoskeleton has been recognized as a promising addition to the methods of treating glaucoma. However, all known actin-targeting small molecules are natural products or close analogues with complex chemical structures, which are difficult to synthesize or to gather from natural sources such as marine sponges. As a consequence there is a need for novel chemical structures as lead compounds targeting actin.

Different computational approaches aimed at identifying actin-binding compounds are used in this thesis. The first method is virtual screening, the most commonly used approach to hit identification in computer-aided drug discovery, which is based on docking and ranking a large collection of compounds. Here, a library of small molecules obtained by decomposing natural products is screened since natural products have been selected by natural evolution to interact with biological targets such as proteins. Therefore, a natural product-inspired library is expected to be enriched in biological activity. The hits obtained by the virtual screening campaign are validated using explicit water molecular dynamics (MD) simulations. The binding mode of compounds with a high binding affinity is expected to be stable in multiple MD runs on a nanosecond time scale. Several molecules with a promising binding behavior in the MD simulations are suggested for chemical synthesis and *in vitro* binding assays. The stability of one hit is significantly improved by *in silico* optimization based on the most populated structural clusters obtained from the MD runs.

De novo design is employed as a second approach to identify novel actin-targeting small molecules. The unique thiazolidinone ring of latrunculin is used as an anchor for the attachment of additional fragments to exploit the optimal hydrogen bond forming properties of this moiety for the actin binding site.

Lastly, MD simulations are carried out to elucidate the mechanism by which latrunculin inhibits actin polymerization. It is shown that latrunculin prevents conformational changes necessary to the transition from monomeric to filamentous actin.

Zusammenfassung

In der vorliegenden Dissertation werden computergestützte Methoden zur Untersuchung des Proteins Aktin angewandt, ein wichtiger Bestandteil des Zytoskeletts. Aktin stellt ein interessantes Drug-Target (Zielmolekül) für die Behandlung von Krebs dar, da die Organisation und der Polymerisationsgrad des Aktinnetzwerks in Tumorzellen bösartig verändert ist. Des Weiteren wurde die lokale Zerstörung des Aktinzytoskeletts als vielversprechende neue Methode zur Behandlung von grünem Star (Glaukom) erkannt. Bei allen bekannten Aktin-bindenden Molekülen handelt es sich jedoch um Naturstoffe oder verwandte Analoga mit komplexer chemischer Struktur, die schwer zu synthetisieren oder aus natürlichen Quellen wie marinen Schwämmen zu gewinnen sind. Aus diesem Grund existiert ein Bedarf an neuartigen chemischen Strukturen für Leitwirkstoffe, die auf Aktin abzielen.

Verschiedene computergestützte Ansätze werden in dieser Dissertation zur Identifikation neuer Aktin-bindender Verbindungen benutzt. Bei dem ersten Ansatz handelt es sich um virtuelles Screening, die am häufigsten verwendete Methode zur Hit-Identifikation im computergestützten Wirkstoffdesign, die auf dem Docking und Ranking von großen Molekülbibliotheken basiert. In dieser Arbeit wird eine Bibliothek aus Molekülen gescreent, die durch Zerlegung von Naturstoffen erhalten wurden, da Naturstoffe durch natürliche Evolution entsprechend ihrer Fähigkeit zur Interaktion mit Biomolekülen wie Proteinen selektiert wurden. Aus diesem Grunde ist eine Naturstoff-basierte Bibliothek vermutlich angereichert an Strukturen mit biologischer Aktivität. Die durch virtuelles Screening erhaltenen Hits werden mit Hilfe von Moleküldynamik- (MD-) Simulationen unter expliziter Berücksichtigung des Lösungsmittels validiert. Für Verbindungen mit hoher Bindungsaffinität wird erwartet, dass die Bindungsmoden im Nanosekunden-Maßstab stabil sind. Einige Moleküle mit vielversprechendem Bindeverhalten in den MD-Simulationen werden für chemische Synthese und Untersuchung der Bindungsaffinität durch *in vitro* Bindungsassays vorgeschlagen. Die Stabilität eines Hits konnte durch computergestützte Optimierung stark verbessert werden, die auf den am meisten populierten strukturellen Clustern der MD-Simulationen basiert.

Bei der zweiten angewandten Methode zur Identifikation neuartiger Aktin-bindender Moleküle handelt es sich um *De Novo*-Design. Der einzigartige Thiazolidinon-Ring des Aktin-Inhibitors Latrunculin wird als Ankerfragment zur Addition weiterer Fragmente verwendet, um seine optimale Fähigkeit zur Ausbildung von Wasserstoffbrückenbindungen in der Aktin-Bindetasche

auszunutzen.

Schließlich werden MD-Simulationen zur Aufklärung des Mechanismus durchgeführt, mit dem Latrunculin die Polymerisation von Aktin inhibiert. Es wird gezeigt, dass Latrunculin Konformationsänderungen verhindert, die notwendig für den Übergang von monomerem zu polymerem Aktin sind.

List of publications

Inhibition of interdomain motion in G-actin by the natural product latrunculin: A molecular dynamics study.
S. Rennebaum, A. Caflisch
[*Proteins*, **2012** , doi: 10.1002/prot.24088]

Targeting G-actin by virtual screening of a natural product-inspired library and validation of hits by molecular dynamics simulations.
S. Rennebaum et al.
[*To be submitted*]

Contents

Chapter 1

Introduction

1.1 The Target: Actin

1.1.1 Actin as a Major Constituent of the Cytoskeleton

The first electron micrographs of cells gave the impression that cells consist of organelles floating freely in a gel-like medium called cytosol. [1] By improved light and electron microscopy techniques it was later discovered that the cytosol is pervaded by a network of filaments, the cytoskeleton. Together cytosol and cytoskeleton make up the cytoplasm of a cell. The cytoskeleton consists of three types of filaments: microtubules, microfilaments, and intermediate filaments. The cytoskeletal filaments differ in size and the type of protein they are composed of. Microtubules consist of the protein tubulin and constitute the thickest cytoskeletal filaments, with an average diameter of 25 nm. Intermediate filaments have a diameter of 10 nm and consist of a variety of different proteins, most of which belong to the keratin family. The thinnest filaments are the microfilaments with a diameter of 5-7 nm. They consist of actin, the most abundant protein in many eukaryotic cells. [2,3] As a main constituent of the cytoskeleton, actin is of fundamental importance for various cellular functions such as cell division, morphology, motility, cytokinesis, adhesion, and muscle contraction. Due to its crucial role for the functioning of cells, its amino acid sequence and structure are highly conserved. The monomeric form (G-actin) assembles to the biologically active filamentous structure (F-actin) in the form of double-stranded helices. The dynamic equilibrium between assembly and disassembly leads to a continuous rearrangement of the actin cytoskeleton, a process controlled by a variety of actin-regulating proteins. Examples of actin-binding proteins are profilin and DNase I which both bind to monomers, the filament-capping and -severing gelsolin, and cofilin which severs filaments and accelerates depolymerization at the pointed end. [4]

Actin is involved in cell movement, usually in association with the motor protein myosin. In the contractile apparatus of muscle cells, actin filaments are arranged in parallel rows, and the space in between is filled with myosin filaments, an arrangement called a myofibril. A contraction of the muscle cell occurs when actin and myosin filaments slide past each other, a process caused by conformational changes of myosin driven by ATP. [1]

Interactions of actin and myosin are, however, not only responsible for the contraction of muscles, but also for movements of nonmuscle eukaryotic cells such as the crawling motion of cells across a surface. Moreover, F-actin acts as “tracks” for the transport of organelles and vesicles within eukaryotic cells, which is based on myosin sliding along the actin filaments. [1]

The actin cytoskeleton is manipulated by a wide range of bacterial and viral human pathogens such as *Salmonella*. [5] A variety of different mechanisms exist by which microbial pathogens exploit the actin cytoskeleton of a host cell to promote infection. One example is an increase in cell motility by changing normal actin organization and dynamics, thereby facilitating propagation of the pathogen.

1.1.2 Structure of G-actin

G-actin has a molecular weight of 42 kDa and consists of 375 residues. The three dimensional structure of G-actin is relatively flat. The single polypeptide chain folds around a deep cleft, the nucleotide binding site, into two domains of similar size which are commonly subdivided into four subdomains (see chapter 4, Fig. 1). Based on the characteristics of its fold, actin is allocated to the same superfamily as hexokinases, sugar kinases, Hsp70 proteins, Arp proteins, and the prokaryotic actin-like homologues MreB and ParM. [6]

Subdomain 2 contains the most mobile part of actin, the DNase I binding loop (D-loop), whose coordinates are not resolved in many crystal structures due to its high flexibility. Both N- and C-terminus lie in subdomain 1. Since the G-actin crystal structure was first solved in 1990 [7], more than 80 different crystal structures have been published of G-actin in complex with different nucleotides, actin-binding proteins, or small molecules such as latrunculin. In all G-actin crystal structures the two domains are arranged in a propeller-like twist relative to each other.

1.1.3 Filament structure and dynamics

The two ends of an actin filament exhibit different structural and dynamical properties, which has important implications for the dynamic equilibrium between G-actin and F-actin and filament organization. Due to a higher polymerization rate at the barbed or plus (+) end, a net elongation of the filament occurs, while the pointed or minus (-) end has a higher dissociation than association rate causing a net depolymerization and shortening of the filament at the pointed end. This phenomenon is called “treadmilling” because a monomer incorporated into the filament at the barbed end apparently “moves” through the polymer to the pointed end as if it were on a treadmill. Treadmilling constitutes the basic mechanism of cell motility by reorganization of the actin cytoskeleton. [8]

The energy required for actin polymerization is supplied by the ATP molecule bound within the deep cleft between the two domains of G-actin. ATP is hydrolyzed after incorporation of the monomer unit into F-actin with a half time of about 2 s. [9] The subsequent release of the inorganic phosphate is an even slower process with a half time of a few minutes. The slow rates lead to the existence of three structurally and functionally different regions in an actin filament: an ATP region at the growing barbed end, an intermediary ADP-P_i region, and an ADP region at the shrinking pointed end (Fig. 1.1). The nature of the nucleotide affects the conformation of the actin subunit it is bound to, and is an important modulator of the binding affinity of actin-regulating proteins to the filament. ADP-actin has a much lower tendency to polymerize than ATP-actin. Profilin promotes nucleotide exchange of ADP-actin, and the resulting ATP-actin can again be incorporated into an actin filament.

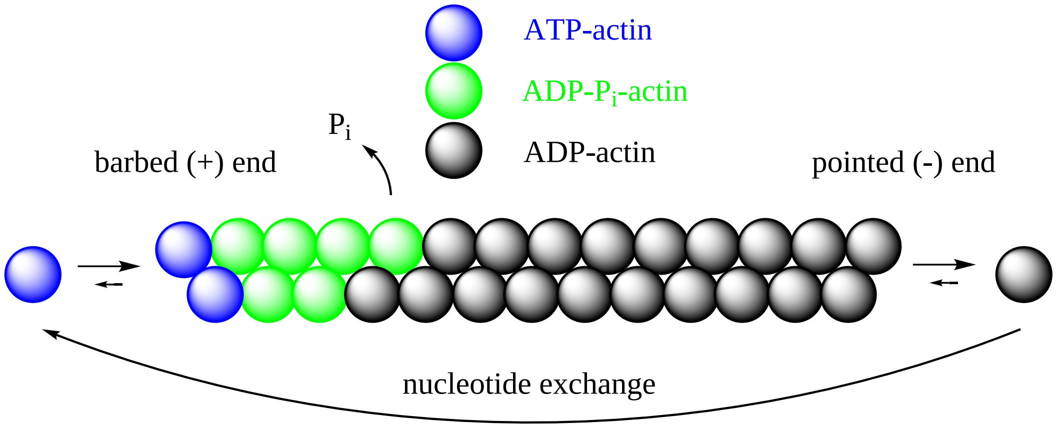


Figure 1.1: Schematic picture of actin filament dynamics. Due to the slow hydrolysis of ATP and even slower release of inorganic phosphate, F-actin consists of an ATP region (blue), an ADP-P_i region (green), and an ADP region (gray).

Due to the difficulties associated with the crystallization of actin filaments, the exact structure of F-actin remains to be determined. [6] The first F-actin model was constructed by Holmes after the solution of the G-actin crystal structure in 1990. [7] The model was based on low-resolution X-ray fiber diffraction data. In the following years, several refined F-actin models were published, based on different techniques such as a directed mutation algorithm [10] and normal mode analysis [11]. An important step towards elucidating the structure of the actin filament was achieved by Oda *et al.* in 2009 [12], who designed a model based on X-ray fiber diffraction data with a resolution of 3.3 Å in the radial direction and 5.6 Å along the equator. The model revealed that a monomer subunit undergoes a significant conformational change when incorporated into the filament, an interdomain rotation of about 20 degrees resulting in a flattening of the monomer subunit. The flattening was confirmed by cryo-EM [13, 14] and a recently published refined F-actin model by Holmes *et al.* [15]. A schematic illustration of the interdomain rotation is shown in Fig. 1.2.

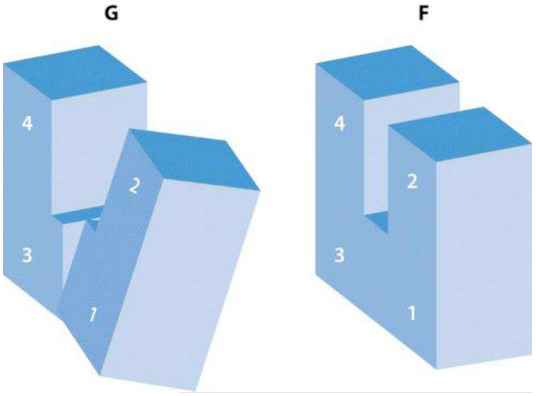


Figure 1.2: Schematic illustration of the conformational change of an actin subunit in the G- to F-actin transition. The twist of the two major domains relative to each other is flattened by an interdomain rotation. The picture is taken from R. Dominguez and K. C. Holmes *Annu. Rev. Biophys.* 2011, 40:169-186.

In chapter 4 the influence of the natural product latrunculin on the interdomain rotation associated with the G- to F-actin transition is investigated.

1.1.4 Actin-targeting Small Molecules as Potential Drugs

Drugs designed to modify the actin cytoskeleton have been recognized as promising therapeutic agents against different diseases such as cancer and glaucoma. [16–20] In tumor cells the molecular organization and composition of the actin cytoskeleton is substantially transformed. As these alterations are directly related to abnormal growth, invasion, and metastasis of malignant cells, actin constitutes a potential target against cancer. Drugs modulating actin polymerization could partially reverse the changes to the actin cytoskeleton in tumor cells and prevent further tumor growth and progression. [21]

The major drawback of any actin-targeting drug that cannot distinguish between normal and malignantly modified actin is the likely occurrence of substantial side effects. However, this argument could also be applied to a few common chemotherapy drugs such as Taxol (paclitaxel). Taxol targets the microtubule cytoskeleton and is unable to distinguish between microtubules in normal and in tumor cells, but is still a useful anticancer agent because malignant cells divide faster and consequently are more sensitive to mitotic inhibition than normal cells. Actin-targeting drugs which accumulate faster in tumor cells could therefore significantly reduce side effects.

A variety of natural products from different sources such as fungus and marine sponges have been shown to target the actin cytoskeleton. [21] Compounds binding to the barbed end are numerous and structurally diverse. Many have a long “tail” connected to a macrolide ring, such as swinholide A and jasplakinolide. The only class of molecules binding to the pointed end are latrunculins, the most potent naturally occurring representative of which is latrunculin A. [22,23] The influence of latrunculin A on the dynamical properties of actin is the focus of chapter 4. Latrunculins have strong anticancer effects, and latrunculins with a reduced binding affinity to actin have been identified which show substantial anti-migratory and anti-invasive activity without being cytotoxic. [24] In addition to their potential as lead structures, actin-targeting natural products are frequently used as molecular probes to dissect cellular pathways. [21]

Due to its ability to reduce intraocular pressure by increasing aqueous humor outflow in higher mammals, latrunculin is a promising lead compound for the treatment of glaucoma. [20] Glaucoma is a disease of the eye in which vision is irreversibly lost, at least partly caused by an increased intraocular pressure damaging the optic nerve. A local treatment with latrunculin has been recognized as a promising antiglaucoma strategy: Latrunculin increases the outflow facility of the eye by disrupting the actin cytoskeleton in the so-called trabecular meshwork, a type of tissue located adjacent to the cornea, thereby reducing pressure on the optic nerve.

1.2 Computational Methods

1.2.1 *De Novo* Drug Design with GANDI

Computational approaches to drug design are valuable tools for the rational design of drugs. The most commonly used method in computer-aided drug discovery is virtual screening, in which pre-existing compounds are docked into a protein structure determined by X-ray crystallography, NMR spectroscopy, or homology modeling. An interesting alternative is the construction of ligands “from scratch”, an approach referred to as *de novo* design which is latin for “from the beginning”. [25–27] It is less commonly used than virtual screening because the design of potent binders which are synthetically accessible is a challenging task. A number of ways have been proposed to deal with the issue of synthetic tractability, such as decomposing existing molecules and recombining the resulting fragments according to a set of rules derived from organic synthesis. [28]

The main advantage of *de novo* design over virtual screening is the much larger chemical space accessible. Still, it is not feasible to evaluate all theoretically possible compounds because chemical space is virtually infinite. [25]

An example of a *de novo* design program is GANDI (Genetic Algorithm-Based de Novo Design of Inhibitors). [27] GANDI employs a genetic algorithm, i.e., an algorithm which mimics Darwin’s theory of natural evolution. A population is evolved by mutation, cross-over, and selection, and both the parent and children population compete for survival. In GANDI the genetic material of an individual consists of one chromosome with as many genes as fragments to connect. A specific gene value corresponds to one docked fragment position and constitutes the value to be optimized by the genetic algorithm. Several populations are evolved simultaneously in GANDI, an approach called a “parallel” or an “island” genetic algorithm. The different populations exchange genetic material after a specified number of steps. The newly introduced genetic material helps to prevent premature convergence by allowing the algorithm to escape a local minimum.

The docked fragments are connected by linker molecules. In contrast to the fragments, the linkers are chosen and connected to the fragments by a random tabu search rather than a genetic algorithm. The reason for decoupling the treatment of the fragments and the linkers is that generally only few linkers are able to link two fragments. Including the choice of linkers in the genetic algorithm would therefore increase computation time by exploring unfeasible regions of search space.

Several scoring functions such as the force field energy, the 3D and 2D structural similarity, and a pharmacophore scoring function are simultaneously optimized, an approach referred to as multiobjective optimization. The individual scores can either be combined by a weighted-sum or a Pareto-based method. The weighted-sum approach uses a linear combination of scoring function terms:

$$S_{\mathrm{total}} = w_{\mathrm{ff}} E_{\mathrm{ff}} - w_{\mathrm{2D}} \mathrm{Sim}_{\mathrm{2D}} - w_{\mathrm{3D}} \mathrm{Sim}_{\mathrm{3D}} - w_{\mathrm{PH4}} \mathrm{Sim}_{\mathrm{PH4}} \tag{1.1}$$

where S_{total} is the total score, E_{ff} the force field energy, $\mathrm{Sim}_{\mathrm{2D}}$ and $\mathrm{Sim}_{\mathrm{3D}}$ the 2D and 3D similarity to a known ligand, and $\mathrm{Sim}_{\mathrm{PH4}}$ the pharmacophore score. The weights for the scoring function terms are w_{ff}, w_{2D}, w_{3D}, and w_{PH4}, respectively. The 2D structural similarity is evaluated by the Tanimoto coefficient between 2D fingerprints.

Scaffold hopping can be promoted by using a high weight for the 3D similarity term, which considers the spatial distribution of atoms of the binding mode of the known ligand, but not its covalent structure. In the Pareto-based scoring an individual A dominates over another individual B if the scores of all scoring functions of A are equal or lower (at least one) than the corresponding scores of B. The advantage of a Pareto-based scoring is that it does not require the tuning of any scoring function weights.

1.2.2 Molecular Dynamics Simulations

Introduction

Molecular dynamics (MD) simulations are a powerful tool to study structural and dynamical properties of biomolecules. In contrast to more accurate, but computationally expensive quantum mechanical methods, electrons are not treated explicitly in MD simulations. Instead, an atomistic description is often employed which, in conjunction with the use of classical mechanics, renders MD simulations applicable to large biomolecules such as proteins or nucleic acids.

The first MD simulation was published in 1977 by McCammon *et al.* of the bovine pancreatic trypsin inhibitor (BPTI) and was 9 ps long. [29] Today simulations in the microsecond time scale are feasible due to substantial improvements in computer performance and the development of more efficient algorithms. MD simulations complement experimental methods and can be used to support and help to understand experimental results. In cases where the desired quantity is not accessible by experimental methods, they offer an alternative way to investigate the biological system or process of interest.

An MD simulation can be regarded as a virtual experiment where the atoms of a system are allowed to propagate for a specified amount of time, and time-dependent processes are monitored. A trajectory describing the propagation of the system over the course of time is obtained by numerically integrating Newton's equation of motion

$$\mathbf{F}_i = -\frac{\partial U(\mathbf{x})}{\partial \mathbf{x}_i} = m_i \ddot{\mathbf{x}}_i \tag{1.2}$$

where U is the potential energy of the system, and $\mathbf{F}_i$, $\mathbf{x}_i$, and m_i are the force, position, and mass of atom i, respectively.

An important parameter to set in an MD simulation is the integration time step. The time step has to be shorter than the highest atomic oscillation rate to be monitored. However, using smaller time steps leads to higher computational costs. Chemical bonds involving hydrogens have the highest oscillation rate, necessitating a time step of appr. one femtosecond. A frequently used approximation is to fix the length of all bonds involving hydrogens, allowing for a time step of about 2 fs to be used.

For biologically relevant time scales only empirical force fields are fast enough for MD simulations of large biological systems such as proteins and nucleic acids. Empirical potential energy functions, called molecular mechanics force fields, are used to describe the relationship between structure and energy. Force fields constitute a compromise between accuracy and computational cost. Different force field such as CHARMM [30], AMBER [31], GROMOS [32], and OPLS [33] can differ both in mathematical form and set of empirical parameters. Because electrons are not treated explicitly, bond formation and breaking processes cannot be simulated by MD. A further approximation used in MD simulations with fixed-charge force fields is neglecting charge polarization. In the CHARMM force field, the potential energy is approximated by a sum of covalent and non-covalent energy contributions, where the former is divided into a bond length, bond angle, dihedral angle, and improper dihedral angle term, whereas the latter is described by a sum of a Coulombic potential to incorporate electrostatic interactions and the Lennard-Jones potential to approximate van der Waals interactions:

$$U(R) = \sum_{bonds} K_b(b-b_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 + \sum_{dihedrals} K_\chi(1+\cos(n\chi-\delta))$$

$$+\sum_{impropers} K_{imp}(\phi-\phi_0)^2+\sum_{nonbond}\left(\epsilon_{ij}\left[\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{12}-2\left(\frac{R_{ij}^{min}}{r_{ij}}\right)^{6}\right]+\frac{q_i q_j}{4\pi\epsilon r_{ij}}\right) \tag{1.3}$$

where K_b, K_θ, K_χ, and K_{imp} are the force constants for the bond, angle, dihedral, and improper dihedral term, and b_0, θ_0, and ϕ_0 the equilibrium values for the bond length b, angle θ, and improper dihedral ϕ. In the dihedral angle term, n is the multiplicity, θ the dihedral angle and δ the phase shift. In the non-covalent terms, R_{ij}^{min} is the distance between atoms i and j where the Lennard-Jones potential is minimal, ϵ_{ij} is the depth of the potential energy well, q_i and q_j the charges of atoms i and j, and r_{ij} the distance between the two atoms.

Before an MD simulation can be carried out, the atom positions have to be energy-minimized. In an X-ray structure some bonds can be stretched or shortened, and large forces would act on the atoms if a simulation were started directly from such a structure, leading to an unphysical acceleration of the atoms. To prevent this behavior, the net force acting on an atom, i.e., the gradient of the potential energy with regard to the atomic positions, is brought to zero by performing an energy minimization

Force Field Parameters for Druglike Molecules

One of the most commonly used empirical force fields is the CHARMM force field mentioned above. Until recently, druglike molecules could not be simulated accurately due to the lack of appropriate parameters in the CHARMM force field. In 2010 this issue was addressed by the introduction of the CHARMM General Force Field (CGenFF). [34] The CGenFF contains parameters for a wide range of chemical structures and functionalities commonly present in druglike molecules such as heterocyclic moieties. It is fully compatible with the CHARMM22 force field.

For this thesis CGenFF parameters for putative actin inhibitors were assigned using the convenient ParamChem server (https://www.paramchem.org) which assigns CGenFF parameters for an uploaded molecule.

Analysis of Molecular Dynamics Trajectories

Root Mean Square Deviation

The root mean square deviation (RMSD) between two conformations of a molecule is the most commonly used measure of structural similarity, e.g. between a protein structure taken from an MD snapshot and the corresponding crystal structure used as reference. Before computing the RMSD, the structures to be compared have to be superimposed. For N atoms the RMSD is defined as

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left[(x_i - x_{ref})^2 + (y_i - y_{ref})^2 + (z_i - z_{ref})^2\right]} \tag{1.4}$$

where x_i, y_i, and z_i are the coordinates of atom i in one structure, and x_{ref}, y_{ref}, and z_{ref} the coordinates of atom i in the reference structure. Time series of RMSD from the crystal structure are a useful measure of the overall structural stability of a protein over the course of an MD simulation.

Root Mean Square Fluctuation

The RMSF constitutes a measure of atomic fluctuations and mobility observed in an MD simulation and is defined as

$$RMSF_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left[(x_i(j) - x_{ref})^2 + (y_i(j) - y_{ref})^2 + (z_i(j) - z_{ref})^2\right]} \tag{1.5}$$

for atom i, where the sum is over all n snapshots in a trajectory. The coordinates of atom i in snapshot j are designated as $x_i(j)$, $y_i(j)$, and $z_i(j)$. In contrast to the RMSD where atomic fluctuations are averaged over all atoms of a molecule for a specific point in time, the root mean square fluctuation (RMSF) averages positional deviations of an atom over the course of a simulation, yielding one RMSF value for each atom.

Motional Correlations of Atomic Fluctuations

Many of the biological functions proteins are involved in are related to correlated motions of specific parts of the protein. The extent of correlation between atomic displacements can be quantified by the normalized fluctuation correlation C_{ij} between atom pairs averaged over the snapshots of an MD simulation, an approach called dynamic cross correlation (DCC). [35,36] C_{ij} is defined as the normalized covariance between two atom position vectors, i.e., the covariance between the vectors divided by the product of their standard deviations:

$$C_{ij} = \frac{\overline{(r_i - \overline{r_i})(r_j - \overline{r_j})}}{\sqrt{\overline{(r_i^2 - \overline{r_i^2})(r_j^2 - \overline{r_j^2})}}} \tag{1.6}$$

where r_i and r_j are the position vectors of atoms i and j, respectively. It is also possible to examine the motional correlation between entire residues. The normalized fluctuation correlation ranges from $C_{ij} = 1$ for full correlation, i.e., for atomic fluctuations with the same period and same phase, through $C_{ij} = 0$ for no correlation, to $C_{ij} = -1$ for fully anticorrelated displacements, i.e. for motions with the same period and opposite phase. The DCC method returns a matrix of all pairwise normalized fluctuation correlations C_{ij} which can be visualized in form of a cross correlation map such as Fig. 4 in chapter 4. Since the diagonal represents the motional correlation between position vectors of the same atoms, the diagonal elements of the matrix have to indicate full correlation. Normalized fluctuation correlations strongly depend on the choice of atoms and reference frame for alignment prior to the correlation analysis. Note that only linear correlations are captured by the DCC approach.

A second method to compute motional correlations is the linear mutual information (LMI) approach. [36–38] The LMI algorithm is based on a different principle than the DCC method. The LMI method measures how much information about a random variable A is contained in a random variable B, i.e., how much information is shared by the two variables. The mutual information between two random variables A and B is high when the uncertainty of B is reduced by observing A. The LMI between the motions of two atoms i and j is calculated according to

$$I_{lin}(x_i, x_j) = \frac{1}{2}(ln|C_i| + ln|C_j| - ln|C_{ij}|) \tag{1.7}$$

where C_{ij} is the pair-wise covariance matrix, C_i and C_j are the marginal covariance matrices, and x_i and x_j are the deviation of the positional fluctuations of atoms i and j from the mean ($x_i = r_i - \overline{r_i}$ and $x_j = r_j - \overline{r_j}$, respectively). LMI values vary from 0 indicating lack of any correlation to +1, which is complete correlation between atomic displacements. Anticorrelation is not captured by the mutual information measure.

The LMI method is computationally more expensive than the DCC approach, but the former has the important advantage that it is able to capture correlations between perpendicular motions. As the DCC method, the LMI approach is restricted to linear correlations, but a more general MI (mutual information) algorithm exists which is able to detect nonlinear correlations as well.

1.3 Structure of the Thesis

The identification of novel actin inhibitors is the focus of chapters 2 and 3. While virtual screening and MD simulations are employed in chapter 2, *de novo* design is carried out in chapter 3 to design truly novel actin inhibitors. In chapter 4 the influence of the natural product latrunculin on the dynamic properties of G-actin is investigated by explicit water MD simulations, elucidating the mode of action by which latrunculin prevents the polymerization of actin. A summary of the results presented in this thesis and an outlook is given in the Conclusion section.

Bibliography

[1] Campbell NA, Reece JB. Biology. San Francisco (CA): Pearson/Benjamin Cummings. 2008. 8th edition.

[2] Schmidt A, Hall MN. Signaling to the actin cytoskeleton. Annu Rev Cell Dev Biol 1998; 14:305–338.

[3] Pollard TD, Cooper JA. Actin, a central player in cell shape and movement. Science 2009; 326:1208–1212.

[4] Uribe R, Jay D. A review of actin binding proteins: new perspectives. Mol Biol Reports 2009;36:121–125.

[5] Dramsi S, Cossart P. Intracellular pathogens and the actin cytoskeleton. Annu Rev Cell Dev Biol 1998;14:137–166.

[6] Dominguez R, Holmes KC. Actin: Structure and function. Annu Rev Biophys 2011;40:169–186.

[7] Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC. Atomic structure of the actin: DNase I complex. Nature 1990;347:37–44.

[8] Wanger M, Keiser T, Neuhaus JM, Wegner A. The actin treadmill. Canad J Biochem Cell Biol 1985;63:414–421.

[9] Murakami K, Yasunaga T, Noguchi TQ, Gomibuchi Y, Ngo KX, Uyeda TQ, Wakabayashi T. Structural basis for actin assembly, activation of ATP hydrolysis, and delayed phosphate release. Cell 2010;143:275–287.

[10] Michael L, Popp D, Holmes KC. Refinement of the F-actin model against X-ray fiber diffraction data by the use of a directed mutation algorithm. J Mol Biol 1993;234:826–836.

[11] Tirion MM, ben-Avraham D, Lorenz M, Holmes KC. Normal modes as refinement parameters for the F-actin model. Biophys J 1995;68:5–12.

[12] Oda T, Iwasa M, Aihara T, Maéda Y, Narita A. The nature of the globular- to fibrous-actin transition. Nature 2009;457:441–445.

[13] Fujii T, Iwane AH, Yanagida T, Namba K. Direct visualization of secondary structures of F-actin by electron cryomicroscopy. Nature 2010;467:724–728.

[14] Murakami K, Yasunaga T, Noguchi TQ, Gomibuchi Y, Ngo KX, Uyeda TQ, Wakabayashi T. Structural basis for actin assembly, activation of ATP hydrolysis, and delayed phosphate release. Cell 2010;143:275–287.

[15] Splettstoesser T, Holmes KC, Noé F, Smith JC. Structural modeling and molecular dynamics simulation of the actin filament. Proteins: Struct Func Bioinf 2011;79:2033–2043.

[16] Jordan MA, Wilson L. Microtubules and actin filaments: Dynamic targets for cancer chemotherapy. Curr Opin Cell Biol 1998;10:123–130.

[17] Janmey PA, Chaponnier C. Medical aspects of the actin cytoskeleton. Curr Opin Cell Biol 1995;7:111–117.

[18] Giganti A, Friederich E. The actin cytoskeleton as a therapeutic target: state of the art and future directions. Prog Cell Cycle Res 2003;5:511–525.

[19] Lambrechts A, Van Troys M, Ampe C. The actin cytoskeleton in normal and pathological cell motility. Int J Biochem Cell Biol 2004;36:1890–1909.

[20] Peterson JA, Tian B, Geiger B, Kaufman PL. Effect of latrunculin B on outflow facility in monkeys. Exp Eye Res 2000;70:307–313.

[21] Allingham J, Klenchin V, Rayment I. Actin-targeting natural products: structures, properties and mechanisms of action. Cell Mol Life Sci 2006;63:2119–2134.

[22] Nèeman I, Fishelson L, Kashman Y. Isolation of a new toxin from the sponge *Latrunculia magnifica* in the Gulf of Aquaba (Red Sea). Marine Biol 1975;30:293–296.

[23] Morton WM, Ayscough KR, McLaughlin PJ. Latrunculin alters the actin-monomer subunit interface to prevent polymerization. Nat Cell Biol 2000;2:376–378.

[24] Khanfar MA, Youssef DTA, El Sayed KA. Semisynthetic latrunculin derivatives as inhibitors of metastatic breast cancer: Biological evaluations, preliminary structure-activity relationship and molecular modeling studies. ChemMedChem 2010;5:274–285.

[25] Schneider G, Fechner U. Computer-based *de novo* design of drug-like molecules. Nature Rev Drug Discov 2005;4:649–663.

[26] Mauser H, Guba W. Recent developments in *de novo* design and scaffold hopping. Curr Opin Drug Discov Devel 2008;11:365–374.

[27] Dey F, Caflisch A. Fragment-based *de Novo* ligand design by multiobjective evolutionary optimization. J Chem Inf Model 2008;48:679–690.

[28] Lewell X, Judd D, Watson S, Hann M. RECAP-Retrosynthetic Combinatorial Analysis Procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. J Chem Inf Model 1998;38:511–522.

[29] McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267:585–590.

[30] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.

[31] Ponder JW, Case DA. Force fields for protein simulations. Adv Prot Chem 2003;66:27–85.

[32] van Gunsteren WF, Daura X, Mark AE. GROMOS Force Field. John Wiley & Sons, Ltd. 2002.

[33] Jorgensen WL, Tirado-Rives J. The OPLS force field for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 1988;110:1657–1666.

[34] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 2010;31:671–690.

[35] McCammon JA, Harvey S. Dynamics of proteins and nucleic acids. Cambridge: Cambridge University Press. 1987.

[36] Seeber M, Felline A, Raimondi F, Muff S, Friedman R, Rao F, Caflisch A, Fanelli F. Wordom: A user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. J Comput Chem 2011;32:1183–1194.

[37] Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E 2004; 69:066138–066154.

[38] Lange OF, Grubmüller H. Generalized correlation for biomolecular dynamics. Proteins: Struct Func Bioinf 2006;62:1053–1061.

Chapter 2

Targeting G-actin by virtual screening of a natural product-inspired library and hit validation by molecular dynamics simulations

Abstract

Natural products constitute the part of chemical space selected by natural evolution to interact with a wide range of biological targets with high binding affinity and selectivity. Compared to standard compound collections, a natural product-inspired library is expected to be enriched in molecules possessing the chemical properties necessary for binding to proteins, thereby facilitating the discovery of novel therapeutic agents. Here, we present a virtual screening campaign of a natural product-derived library to identify compounds targeting actin. Actin inhibitors are promising candidates for the development of novel anticancer and antiglaucoma drugs. To obtain a more accurate ranking and validate the hits obtained by high-throughput docking and reranking by consensus scoring, the stability of the binding modes of selected compounds is examined by explicit water molecular dynamics simulations. Several molecules with a promising binding behavior in the simulations are suggested for chemical synthesis and *in vitro* binding assays. Addition of only two heavy atoms to one of the hits significantly improved the stability of the most populated binding mode, suggesting a higher binding affinity of the derivative compared to the original compound.

Introduction

Natural products (NPs) constitute a valuable source of novel chemical structures in the search for new therapeutic agents. [1,2] Under evolutionary pressure they have been selected by nature for a high binding affinity to specific biological targets and the ability to induce a biological effect, such as activation or inhibition of a protein. Examples of NPs used as drugs include the anticancer agent paclitaxel (Taxol), which is isolated from the bark of the Pacific yew tree and targets microtubules [3], as well as the fungal metabolite cyclosporine which suppresses the immune response after organ transplants [4].

A number of structurally diverse NPs have been identified capable of interfering with the actin cytoskeleton. [5] Actin is involved in many cellular processes such as motility, division, cytokinesis, and muscle contraction. [6] The reversible assembly of monomeric actin (G-actin) to filamentous actin (F-actin) is controlled by a variety of actin-binding proteins such as profilin, gelsolin, and DNase I. In cancer cells, the F-actin morphology is substantially transformed, making actin a potential drug target against tumor cell growth and metastasis. [7–10] Moreover, disruption of actin filaments has been shown to reduce intraocular pressure in monkeys, which may be useful as treatment of glaucoma. [11] Actin-binding NPs can either stabilize actin filaments (e.g. phalloidin, jasplakinolide) or depolymerize them (e.g. swinholide A, aplyronine A), the latter either by actively disrupting the filaments or by monomer sequestration. [5] Crystal structures show most actin-targeting NPs binding to the hydrophobic cleft at the so-called barbed end of actin between subdomains 1 and 3. [5] As of today only one class of actin inhibitors targeting the pointed end above the nucleotide binding site is known, the latrunculins, which are macrolide toxins containing a unique thiazolidinone ring. They are produced by sponges in the Red Sea as chemical defense against predators. [12,13]

Since latrunculin as well as other actin-binding NPs are difficult to synthesize or extract from natural sources such as marine sponges, there is a demand for novel chemical scaffolds for the development of actin-targeting anticancer and antiglaucoma drugs. To address this need, we report a high-throughput docking campaign targeting the latrunculin binding site of actin. A library of fragments obtained by decomposing a collection of NPs is used in the virtual screening since NP-derived fragments are expected to possess chemical properties facilitating the binding to biological targets, but are less complex than the original molecules. The decomposition is carried out according to a set of rules derived from organic synthesis, allowing a coupling of promising fragments to synthetically accessible molecules at a later stage of the project.

Since consensus scoring has been shown to yield higher hit rates than the use of only a single scoring function [14,15], the poses are ranked according to the median rank of different scoring functions: the SEED [16,17] total and electrostatic interaction energy, the CHARMm [18] total interaction energy and van der Waals efficiency, the number of rotatable bonds, and the ligand strain. Explicit solvent molecular dynamics (MD) simulations are performed to investigate the binding behavior of the hits found in the virtual screening campaign at an atomistic level of detail. The unbinding times calculated by a single-exponential fit of the cumulative distribution of unbinding events suggest a micromolar activity for most compounds. [19] Ligands with a stable binding mode and high unbinding time in MD simulations are selected for chemical synthesis and experimental determination of binding affinity. Furthermore, an *in silico* optimization of a promising compound was carried out using the most populated binding modes obtained by clustering of the MD trajectories. Addition of only two heavy atoms improved both hydrophobic contacts and the stability of hydrogen bonds to the binding site and lowered the number of spontaneous unbinding events, indicating a higher binding affinity to actin for the optimized compound.

Materials and Methods

Protein preparation

The coordinates of G-actin in complex with latrunculin A [13] were downloaded from the PDB database (PDB ID: 1ESV). The inhibitor and all water molecules were removed. As residues 1-5 are missing in the X-ray structure, the -$COCH_3$ group was added to the N-terminal Thr6. The C-terminus was considered negatively charged.

Due to the high flexibility of the DNase I binding loop (D-loop), the coordinates of its residues are not present in the PDB file (residues 40-50). As in previously published MD studies of G-actin, [20–23] initial D-loop coordinates were obtained from an actin crystal structure with the D-loop residues resolved. Using SWISS-MODEL, [24] the PDB structure 3DAW was selected as template for the two following reasons. First, the 1ESV coordinate set is an ATP-bound actin structure, which is also the case for 3DAW. Second, in 3DAW the protein used for co-crystallization (which is a domain of twinfilin) binds between the actin subdomains 1 and 3 which are located far away from the D-loop, and thus the presence of the twinfilin domain does not directly influence the conformation of the D-loop. Similar to the procedure described in [21], after fitting the crystal structure to the template using the C_α atoms present in both structures, the missing coordinates of the D-loop were taken from the template and inserted into the 1ESV structure. To relax elongated bonds after the coordinate transfer, an energy minimization of 100 steps of the steepest descent (SD) and 1000 steps of the adopted basis Newton-Raphson algorithm was carried out using the program CHARMM [25, 26] and the CHARMM22 force field [27]. During the minimization a distance-dependent dielectric function of $\epsilon = 4r$ was used, and all coordinates present in the X-ray structure were kept fixed. To reproduce neutral pH conditions, the side chains of aspartates and glutamates were negatively charged, those of lysines and arginines were positively charged, and histidines were considered neutral.

Preparation of the library of natural product-derived fragments

A library of natural products was assembled by downloading the NP subsets of the August 2009 versions of the ZINC library [28], the Crystallography Open Database (http://www.crystallography.net), the Human Metabolome Database (http://www.hmdb.ca), MicroSource Discovery (http://www.msdiscovery.com), and the Traditional Chinese Medicine Database (http://tcm.cmu.edu.tw). Compounds containing metals were removed, yielding a total of about 140,000 molecules. Babel [29] was used to assign protonation states according to a pH of 7.4, and partial charges were calculated with the MPEOE approach using Witnotp [30]. The compounds were decomposed into molecular fragments according to a set of rules derived from chemical retrosynthesis (Supplementary Material, Fig. S1) similar to the RECAP [31] rules using the programs ReCore [32] and CoLibri [33]. The main difference between the applied rules and the RECAP rules is that not all specified bonds are cut at the same time, but all possible fragment combinations are generated from the original molecules by a successive application of the cleavage rules. This way interesting structural motifs of the natural products are preserved in the resulting fragments. Only fragments with a molecular weight between 50-400 g/mol and up to 10 rotatable bonds were kept.

CHARMm atom types [18] were assigned to the remaining 150,000 fragments with Witnotp, and all fragments were subject to an energy minimization using the CHARMm force field and the program CHARMM [25, 26]. A distance-dependent dielectric function of $\epsilon = 4r$ was used in the minimization.

High-throughput docking

The NP-derived library was docked with version 4.2 of AutoDock [34] using a rigid protein. First, AutoGrid was employed to generate the atom-specific affinity map files. The spacing between two adjacent grid points was 0.375 Å, and the numbers of points in the x, y, and z directions were 76, 62, and 58, respectively. Subsequently, the Lamarckian genetic algorithm of AutoDock, a hybrid genetic algorithm with local search, was run 50 times with different initial seeds to obtain multiple poses, on average 17 poses per compound. The maximum number of energy evaluations was set to 2,750,000 and the maximum number of generations to 27,000. The poses generated by AutoDock were energy-minimized in the rigid protein with CHARMM using the CHARMm force field and a distance-dependent dielectric function of $\epsilon = 4r$.

Consensus scoring

Combining multiple scoring functions for rescoring a set of docked poses, an approach called consensus scoring, has been shown to increase performance in hit rates compared to using only a single scoring function. [14, 15] The energy functions used for consensus scoring in this study are:

1) The SEED [16,17] protein-compound total interaction energy.
2) The SEED protein-compound electrostatic interaction energy.
3) The CHARMm [18] protein-compound total interaction energy.
4) The CHARMm protein-compound van der Waals interaction efficiency, i.e., the van der Waals interaction energy divided by the number of heavy atoms.
5) The number of rotatable bonds of the ligand.
6) The conformational strain of the ligand in the docked conformation, defined as the difference in intra-ligand energy in the protein-bound and -unbound conformation.

The SEED interaction energies were calculated using the SEED energy evaluation mode. The SEED scoring function includes electrostatic interactions as well as receptor and fragment desolvations calculated by an efficient numerical implementation of the generalized Born approach. [16,17]

Ranking the poses according to the number of rotatable bonds is a simple approximation to incorporate the entropic penalty of the ligand upon binding. For ligands with many rotatable bonds the entropic cost due to loss of rotational degrees of freedom upon binding is higher than for more rigid ligands. Note that in this simple approximation, a linear chain with six rotatable bonds has the same estimated penalty as benzene with six -OCH_3 substituents, for example, even though the latter will have a much lower entropic penalty than the former upon binding. The ranks according to the number of rotatable bonds were assigned by giving all n compounds with no rotatable bonds a rank of 1. To the m compounds with one rotatable bond, a rank of $n + 1$ was assigned, to those with two rotatable bonds a rank of $n + m + 1$, and so forth.

The ligand strain was calculated by minimizing the ligand conformation obtained by docking first within the rigid protein and again after deletion of the protein atoms, using a distance-dependent dielectric function of $\epsilon(r)=4r$ in both cases, and calculating the difference between the two intramolecular energies of the ligand. Ligands adopting a high-energy conformation are less likely to have high binding affinities. Note that the strain is calculated with respect to the nearest local minimum of the unbound conformation.

After an independent ranking by each energy function, the poses were ranked according to the median rank since the median is less sensitive to outliers compared to the average. [35] The poses were ranked by the median rank of functions 1-4, 1-5, and 1-6, since visual examination of the 400 top-ranking molecules revealed no distinct difference in quality of the binding poses for the three combinations of energy functions. Including the AMBER-based [36] AutoDock scoring function did not yield a higher number of reasonable binding poses among the 400 top-ranking compounds. The AMBER and CHARMm force field are based on similar principles, therefore using both force fields is partially redundant. In the visual inspection special attention was paid to the number and geometry of hydrogen bonds to the key residues Tyr69, Asp157, Thr186, and Arg210, which are also involved in hydrogen bonds in the actin/latrunculin complex, as well as to van der Waals interactions to side chains of hydrophobic residues in the binding site (e.g. Pro32 and Ile34). Ten compounds were selected according to the median rank of energy functions 1-4, one compound according to 1-5, and four compounds according to 1-6, and their binding behavior was studied in MD simulations.

MD simulations

For each ligand 10-20 independent MD runs were carried out with different initial velocities, starting from the docked conformation obtained by AutoDock. Starting poses for derivatives of the selected compounds (see Results section) were designed by changing the chemical structures using Witnotp and minimizing the resulting molecules in the rigid protein using CHARMM and the CHARMm force field. A distance-dependent dielectric function of $\epsilon = 4r$ was used in the minimization.

For promising compounds with a low number of unbinding events in the first 10 runs and reasonable synthetic feasibility, additional runs were performed, explaining the different number of runs for different compounds. The runs were stopped after 20 ns or before if the distance between the ligand center of mass and the center of mass of the binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210) exceeded 20 Å.

For the MD simulations the protein was immersed in an orthorhombic box of pre-equilibrated water molecules. The size of the box was chosen to have a minimal distance of 13 Å between the boundary and any atom of the protein. VMD [37] was used for setting up the simulation system, while minimization, heating and production runs were performed with NAMD [38] using the CHARMM22 force field and the TIP3P model of water. For the parameters of the ligands the CGenFF force field [39] was used. Periodic boundary conditions were applied, and the particle-mesh Ewald approach [40] was used for the long-range electrostatics. The van der Waals interactions were truncated at a cutoff of 12 Å and a switch function was applied starting at 10 Å. The MD simulations were carried out at constant temperature (310 K) and constant pressure (1 atm) with a time step of 2 fs using the SHAKE algorithm [41] to fix the length of covalent bonds involving hydrogen atoms.

Clustering procedure

To determine the most populated binding modes of compound 42070 in the MD simulations, the trajectories of all individual runs of the ligand were merged and then clustered using Witnotp. First, the non-symmetric heavy atoms of the binding site residues were aligned. Subsequently, the ligand poses were clustered using all non-symmetric heavy atoms of the ligand and a clustering cutoff of 2.5 Å. The option “metric inplace rmsd” was chosen in Witnotp to also take translation and not just different orientations and conformations of the ligand into account. The most populated clusters were visualized for *in silico* optimization of the binding behavior of compound 42070.

Results

The workflow implemented in the *in silico* screening campaign is depicted in Fig. 2.1. After decomposing the library of about 140,000 natural products, fragments with a molecular weight higher than 400 g/mol were discarded to allow a margin for later lead optimization which generally results in an increased molecular weight. Moreover, fragments with a molecular weight below 50 g/mol were considered too small and unspecific and were removed from the library as well. Fragments with more than 10 torsional degrees of freedom were also discarded because with an increasing number of rotatable bonds it becomes more difficult for the docking program to find meaningful poses, since a much larger conformational space has to be sampled.

The decomposition according to a set of rules inferred from organic synthesis (Supplementary Material, Fig. S1) produces building blocks which can be recombined to form synthetically accessible structures. However, a large number of fragments has a molecular weight already within the range of leadlike compounds (Fig. 2.2, top left). Due to their relatively high molecular weight the fragments were considered to be whole molecules and a combination of fragments was not performed at this stage of the project, but is interesting for later lead optimization. Most molecules fulfill Lipinski’s rule of five [42], increasing their potential as candidates for orally bioavailable drugs.

Compounds selected by flexible ligand docking and consensus scoring

After docking the NP-derived library and sorting the poses according to the median ranks of the three combinations of scoring functions (see Materials and Methods section), the 400 top-ranking compounds for each combination were visually examined. Considering the quality of the binding poses, synthetic accessibility, as well as chemical diversity, 15 compounds were selected for MD simulations. The chemical structures of the compounds are shown in Fig. S2 and S3 (Supplementary Material). Despite their reactivity, molecules with an aldehyde group were chosen as well since the aldehyde functionality can be replaced by a bioisostere in later optimization steps. In addition to the molecules chosen from consensus scoring, commercially available or synthetically easily accessible derivatives of the chosen compounds were identified (compounds labeled $42070_{methoxy}$, 42070_{phe}, 42070_{tyr}, $102213_{enantio}$, 106499_{deriv1}, and both enantiomers of 106499_{deriv2}). Derivatives of compound 42070 were designed to improve the binding affinity of the original molecule (42070_{opt1}, 42070_{opt1_deriv1}, 42070_{opt1_deriv2}, 42070_{opt2}, 42070_{opt2_5ring}, 42070_{opt3}, 42070_{opt5}), yielding a total of 29 molecules whose binding behavior to actin was investigated by MD simulations.

Only the combination of six scoring functions including the ligand strain produced a ranking in which latrunculin is among the 400 top-ranking compounds. If it were included in the library, latrunculin would have rank 375 in the consensus scoring of scoring functions 1-6, rank 1535 for functions 1-5, and 1231 for functions 1-4. Nevertheless, all three types of consensus scoring were used for hit identification since visual examination showed no significant differences in the quality of binding poses between the three sets of 400 NP-inspired top-ranking molecules.

As an example, the binding mode of compound 42070 predicted by AutoDock is shown in Fig. 2.4. The aromatic ring is fixed deep in the binding site by a hydrogen bond between the phenolic hydroxyl group and Asp157. According to the predicted binding mode Tyr69 is involved in a hydrogen bond to the ether oxygen which is a rather weak acceptor. Arg210 and the carboxylate group of the ligand participate in a strong electrostatic interaction. The isopropyl group of compound 42070 is located in the vicinity of the side chains of Pro32 and Ile34, indicating the existence of hydrophobic contacts.

Examination of ligand binding behavior by MD simulations

To assess the stability of the binding modes predicted by docking and investigate the binding behavior of the selected compounds, 10-20 independent MD simulations per compound were carried out. The low values of the C_α root mean square deviation (RMSD) of actin from the X-ray structure indicate that the overall structural stability of the protein is preserved in the MD simulations (Fig. 2.3).

An overview of the performed MD runs is presented in Table 2.1. The compounds undergo between zero and 14 spontaneous unbinding events, where an unbinding event is defined as a separation of the ligand center of mass from the center of mass of the binding site larger than 10 Å. Comparable numbers of unbinding events are obtained by applying a distance cutoff of 8 Å or 12 Å. For instance, both a 8 Å and a 10 Å cutoff result in 14 dissociation events for compound 104690, while a 12 Å cutoff yields 13 events. For compound 109412 the corresponding values are 8, 7, and 7 unbinding events for 8 Å, 10 Å and 12 Å. A criterion of 10 Å was chosen to be consistent with a previously published study of ligand unbinding for the FK506 binding protein (FKBP) by Huang and Caflisch. [19] Time series of the distances between the ligands and the binding pocket (Supplementary Material, Fig. S4-S32) indicate that multiple binding modes and dissociation pathways exist for most ligands, in agreement with the binding behavior of the FKBP ligands observed in [19]. Compound 42070_{opt1}, for example, dissociates quickly from the binding site in run 16, remains stable for 20 ns in a number of runs such as run 13, and shows a number of different binding modes in run 7, where it almost unbinds at about 10 ns, but then moves partly back into the binding pocket for a couple of nanoseconds (Supplementary Material, Fig. S8). Compound 104690, however, unbinds quickly in 14 out of 20 runs (Fig. S21), while e.g. compound 105302 remains stable in all runs (Fig. S22). Even though compound 106230 does not unbind in run 10, the binding mode is not stable as indicated by the strong fluctuations (Fig. S24). Only 10 runs were carried out of the spiro compound 34985 as it unbinds 9 out of 10 simulations. The disaccharide 37644 is prone to hydrolytic cleavage by metabolic enzymes and therefore not considered for a probably time-consuming synthesis. Only 10 runs of compounds $106499_{deriv2\_enantio1}$ and $106499_{deriv2\_enantio2}$ were performed because in vitro assays did not show a significant binding affinity (unpublished results).

Unbinding times were extracted for each ligand with at least 5 unbinding events by fitting the cumulative distribution of unbinding events with a single exponential function $f(t) = \exp(-t/\tau_{MD}^{fit})$ (Supplementary Material, Fig. S35). Ideally, τ_{MD}^{fit} should be correlated with the experimentally measured binding affinity of the ligands, as observed in [19]. The unbinding times are determined both by a leave-one-out and a block averaging procedure (see caption of Table 2.1). Note, however, that polarization effects are neglected in the fixed-charge CHARMM force field which has a stronger influence on compounds with a higher charge. Therefore, the unbinding times should not be directly compared between ligands of different charge. The good performance of compounds 105302 and 42783, which did not unbind in any of the MD runs, could be associated with their high charge. The determined unbinding times of 7-76 ns suggest an activity within the micromolar range, an estimation based on a linear fitting of the unbinding times of six ligands of FKBP to their experimentally measured binding energies. [19] The single-exponential behavior observed for the cumulative distributions of unbinding events suggests that one predominant free energy barrier exists, the barrier of ligand unbinding, while the barriers between subbasins of the bound state are much lower, as observed in [19]. The subbasins of the bound state correspond to different binding modes.

The *in vitro* binding affinity and cytotoxicity of a number of compounds listed in Table 2.1 are currently being investigated in the Department of Organic Chemistry at the University of Zürich.

In silico optimization

To further investigate the binding behavior of promising compounds, the MD trajectories were clustered according to different ligand binding modes (see Materials and Methods section) and representatives of the most populated clusters were visualized. The three most populated clusters of compound 42070 have almost equal statistical weights of 12%, 11%, and 10% for a clustering cutoff of 2.5 Å. For a cutoff of 2.0 Å the corresponding values are 10%, 9%, and 7%. The binding mode of the second most populated cluster (cutoff of 2.5 Å) is very similar to the binding mode predicted by docking (Fig. 2.4, bottom right). While the position of the aromatic ring in the most populated cluster is similar to the AutoDock pose, the aliphatic side chain has rotated so that the carboxylate group is involved in hydrogen bonds to both Tyr69 and Arg206. The probably weak hydrogen bond between Tyr69 and the ether group in the docked pose is broken and replaced by a hydrogen bond to the nitrogen-bound hydrogen of the amide group.

Based on the binding modes of the most populated clusters, hydrophobic contacts between the apolar side chains of Pro32, Ile34, and Leu67 and compound 42070 can be reinforced by introducing apolar groups close to the isopropyl moiety. Thus, compound 42070 was chosen for *in silico* lead optimization. Moreover, compound 42070 has only one negative charge, not too many unbinding events (6 out of 20 runs), as well as some MD runs with a nearly perfectly stable binding mode (e.g. runs 2 and 10, see Fig. S6) indicating the existence of stable interactions between the molecule and the binding pocket.

The optimization was focused on improving hydrophobic interactions. Generally the burial of hydrophobic surface area upon ligand binding is the single parameter to correlate with binding affinity, while the contribution of hydrogen bonds depends strongly on desolvation effects. [43] Since the native ligand latrunculin has a large apolar surface area, strong hydrophobic contacts to the actin binding site at the pointed end are likely to be of great importance for a high binding affinity.

The phenolic hydroxyl group of compound 42070 is involved in hydrogen bonds in appr. 70% of the simulation time, formed almost exclusively to Asp157 (Fig. S6, top right and bottom left). As a consequence, the phenyl ring is tightly bound deep in the binding site for most of the simulation time. The aliphatic side chain of compound 42070 is more flexible. Its amide and ether groups form hydrogen bonds in less than 30% of all snapshots. The carboxylate group, however, participates in electrostatic interactions in about 70% of the simulation time, predominantly to Arg210 and to a lesser extent to Arg62 and Arg206 (Fig. S6, bottom right).

Based on these observations and visual examination of the binding modes in the most populated clusters, seven derivatives of compound 42070 were designed and their binding behavior examined by MD simulations (compounds 42070_{opt1}, 42070_{opt1_deriv1}, 42070_{opt1_deriv2}, 42070_{opt2}, 42070_{opt2_5ring}, 42070_{opt3}, and 42070_{opt5}). As some torsional degrees of freedom are fixed by the ring structure, the entropic penalty upon binding is likely to be lower for derivatives 42070_{opt2} and 42070_{opt2_5ring} compared to acyclic structures. In addition, the binding behavior of the commercially available compound $42070_{methoxy}$, as well as that of compounds 42070_{phe} and 42070_{tyr}, was investigated by MD. The latter two molecules can easily be synthesized, and the newly introduced phenylalanine- and tyrosine-derived moieties fit into the binding site after minimization with CHARMM.

The unbinding times of the derivatives of compound 42070 are given in Table 2.1. The most promising lead for actin inhibition is derivative 42070_{opt1}. It has only three unbinding events compared to six events of the parent compound. The time series of the distance between 42070_{opt1} and the binding site show a higher stability of the derivative compared to the parent molecule (Supplementary Material, Fig. S8). The average occupancy of the hydrogen bond between Asp157 and the phenolic hydroxyl group is increased from appr. 65% to 79% of the simulation time, indicating the phenyl ring is more stable in the derivative (green columns in Fig. S6). However, the most significant increase is in side chain stability. The carboxylate group forms an electrostatic interaction to Arg210 in 73% of all snapshots compared to 44% for the original molecule 42070. The NH group of the amide, however, is no longer involved in hydrogen bonds in the derivative due to the steric requirements of the newly introduced methyl group at R^1. Hydrophobic contacts between compound 42070_{opt1} are more stable in comparison to compound 42070, as shown by more stable time series of the distance between the ligand and hydrophobic residues (Supplementary Material, Fig. S33 and S34). The three most populated clusters of compound 42070_{opt1} have statistical weights of 21%, 8%, and 7% for a clustering cutoff of 2.5 Å compared to 12%, 11%, and 10% for compound 42070, indicating that the addition of only two heavy atoms has increased the stability of the most frequently occurring binding mode. The highly occupied hydrogen bond between the phenolic hydroxyl group and Asp157 is lost in the commercially available compound $42070_{methoxy}$ due to the methylation of the OH group, providing a reasonable explanation for the much lower unbinding time of this derivative compared to the original compound 42070. The increased hydrogen bond occupancy of the carboxylate group in 42070_{opt5} compared to 42070 is compensated by a decreased occupancy of the amide group. The unbinding times of compound 42070_{opt5} and 42070 are almost equal, indicating the presence of the methyl group at R^1 is key to improving binding affinity.

Conclusions

A virtual screening campaign of a NP-derived library was carried out to discover novel leads targeting actin. Several interesting structures were selected as candidate ligands after reranking the poses by consensus scoring. Investigating the binding behavior using explicit water MD simulations revealed great differences in the stability of the binding modes. While some molecules unbind after a few nanoseconds in many independent runs, others show a significant degree of stability and are promising candidate ligands for lead development. The unbinding times calculated by a single-exponential fit of the cumulative distribution of unbinding events suggest a micromolar activity for most molecules. Addition of only two heavy atoms to compound 42070 improved the stability of the most populated binding mode significantly by decreasing side chain flexibility due to stronger hydrophobic contacts and hydrogen bonds to the binding site. The high stability of the binding mode of the optimized compound suggests this structure has a high potential as lead candidate. As a next step, the binding affinities of the suggested molecules will be determined experimentally by *in vitro* binding assays (work in progress).

Bibliography

[1] Koch MA, Wittenberg LO, Basu S, Jeyaraj DA, Gourzoulidou E, Reinecke K, Odermatt A, Waldmann H. Compound library development guided by protein structure similarity clustering and natural product structure. Proc Natl Acad Sci USA 2004;101:16721–16726.

[2] Paterson I, Anderson EA. The renaissance of natural products as drug candidates. Science 2005;310:451–453.

[3] Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT. Plant antitumor agents. VI. isolation and structure of taxol, a novel antileukemic and antitumor agent from Taxus brevifolia. J Am Chem Soc 1971;93:2325–2327.

[4] Cohen DJ, Loertscher R, Rubin MF, Tilney NL, Carpenter CB, Strom TB. Cyclosporine: A new immunosuppressive agent for organ transplantation. Ann Int Med 1984;101:667–682.

[5] Allingham J, Klenchin V, Rayment I. Actin-targeting natural products: structures, properties and mechanisms of action. Cell Mol Life Sci 2006;63:2119–2134.

[6] Schmidt A, Hall MN. Signaling to the actin cytoskeleton. Annu Rev Cell Dev Biol 1998; 14:305–338.

[7] Jordan MA, Wilson L. Microtubules and actin filaments: Dynamic targets for cancer chemotherapy. Curr Opin Cell Biol 1998;10:123–130.

[8] Janmey PA, Chaponnier C. Medical aspects of the actin cytoskeleton. Curr Opin Cell Biol 1995;7:111–117.

[9] Giganti A, Friederich E. The actin cytoskeleton as a therapeutic target: state of the art and future directions. Prog Cell Cycle Res 2003;5:511–525.

[10] Lambrechts A, Van Troys M, Ampe C. The actin cytoskeleton in normal and pathological cell motility. Int J Biochem Cell Biol 2004;36:1890–1909.

[11] Peterson JA, Tian B, Geiger B, Kaufman PL. Effect of latrunculin B on outflow facility in monkeys. Exp Eye Res 2000;70:307–313.

[12] Nèeman I, Fishelson L, Kashman Y. Isolation of a new toxin from the sponge *Latrunculia magnifica* in the Gulf of Aquaba (Red Sea). Marine Biol 1975;30:293–296.

[13] Morton WM, Ayscough KR, McLaughlin PJ. Latrunculin alters the actin-monomer subunit interface to prevent polymerization. Nat Cell Biol 2000;2:376–378.

[14] Charifson PS, Corkery JJ, Murcko MA, Walters WP. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. J Med Chem 1999;42:5100–5109.

[15] Clark RD, Strizhev A, Leonard JM, Blake JF, Matthew JB. Consensus scoring for ligand/protein interactions. J Mol Graph Model 2002;20:281–295.

[16] Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. Proteins: Struct Func Bioinf 1999;37:88–105.

[17] Majeux N, Scarsi M, Caflisch A. Efficient electrostatic solvation model for protein-fragment docking. Proteins: Struct Func Bioinf 2001;42:256–268.

[18] Momany FA, Rone R. Validation of the general purpose QUANTA3.2/CHARMm force field. J Comput Chem 1992;13:888–900.

[19] Huang D, Caflisch A. The free energy landscape of small molecule unbinding. PLoS Comput Biol 2011;7:e1002002.

[20] Pfaendtner J, Branduardi D, Parrinello M, Pollard TD, Voth GA. Nucleotide-dependent conformational states of actin. Proc Natl Acad Sci USA 2009;106:12723–12728.

[21] Zheng X, Diraviyam K, Sept D. Nucleotide effects on the structure and dynamics of actin. Biophys J 2007;93:1277–1283.

[22] Dalhaimer P, Pollard TD, Nolen BJ. Nucleotide-mediated conformational changes of monomeric actin and Arp3 studied by molecular dynamics simulations. J Mol Biol 2008; 376:166–183.

[23] Saunders MG, Voth GA. Water molecules in the nucleotide binding cleft of actin: Effects on subunit conformation and implications for ATP hydrolysis. J Mol Biol 2011;413:279–291.

[24] Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. Bioinformatics 2006;22:195–201.

[25] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.

[26] Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. J Comput Chem 2009; 30:1545–1614.

[27] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.

[28] Irwin JJ, Shoichet BK. ZINC - a free database of commercially available compounds for virtual screening. J Chem Inf Model 2005;45:177–182.

[29] O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminf 2011;3:33–47.

[30] Widmer A. Witnotp: A computer program for molecular modeling. 1997; Novartis: Basel.

[31] Lewell X, Judd D, Watson S, Hann M. RECAP-Retrosynthetic Combinatorial Analysis Procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. J Chem Inf Model 1998;38:511–522.

[32] Maass P, Schulz-Gasch T, Stahl M, Rarey M. Recore: a fast and versatile method for scaffold hopping based on small molecule crystal structure conformations. J Chem Inf Model 2007; 47:390–399.

[33] Oloff S, Zhang S, Sukumar N, Breneman C, Tropsha A. Chemometric analysis of ligand receptor complementarity: Identifying complementary ligands based on receptor information (Colibri). J Chem Inf Model 2006;46:844–851.

[34] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a lamarckian genetic algorithm and and empirical binding free energy function. J Comput Chem 1998;19:1639–1662.

[35] Klon AE, Glick M, Davies JW. Combination of a naive bayes classifier with consensus scoring improves enrichment of high-throughput docking results. J Med Chem 2004;47:4356–4359.

[36] Ponder JW, Case DA. Force fields for protein simulations. Adv Prot Chem 2003;66:27–85.

[37] Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. J Mol Graph 1996; 14:33–38.

[38] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005; 26:1781–1802.

[39] Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 2010;31:671–690.

[40] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103:8577–8593.

[41] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equation of motion of a system with constraints: Molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341.

[42] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev 2001;46:3–26.

[43] Bissantz C, Kuhn B, Stahl M. A medicinal chemist's guide to molecular interactions. J Am Chem Soc 2010;53:5061–5084.

Tables and Figures

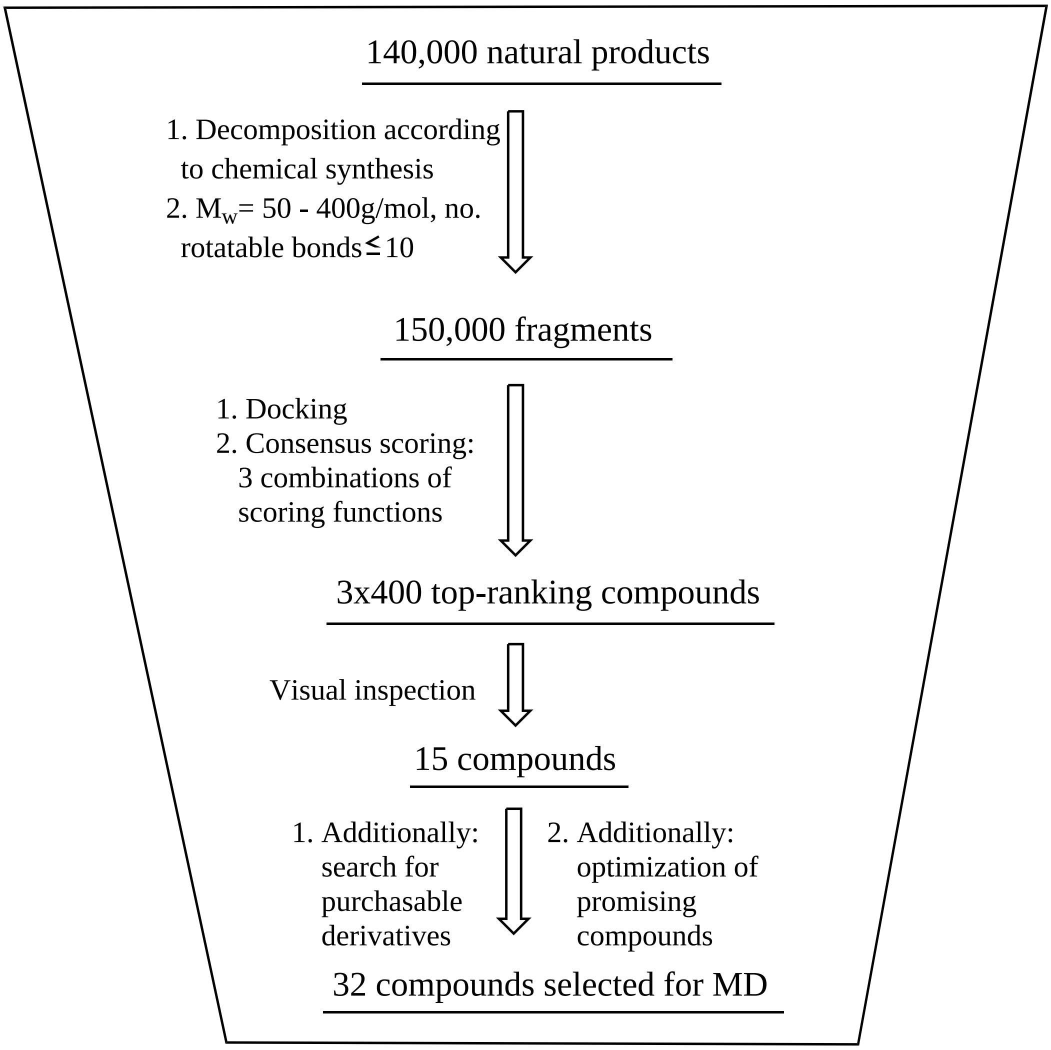


Figure 2.1: Schematic picture of the workflow implemented in the *in silico* screening campaign. Docking of the 150,000 NP fragments was carried out by AutoDock using the 1ESV crystal structure of G-actin as explained in the text.

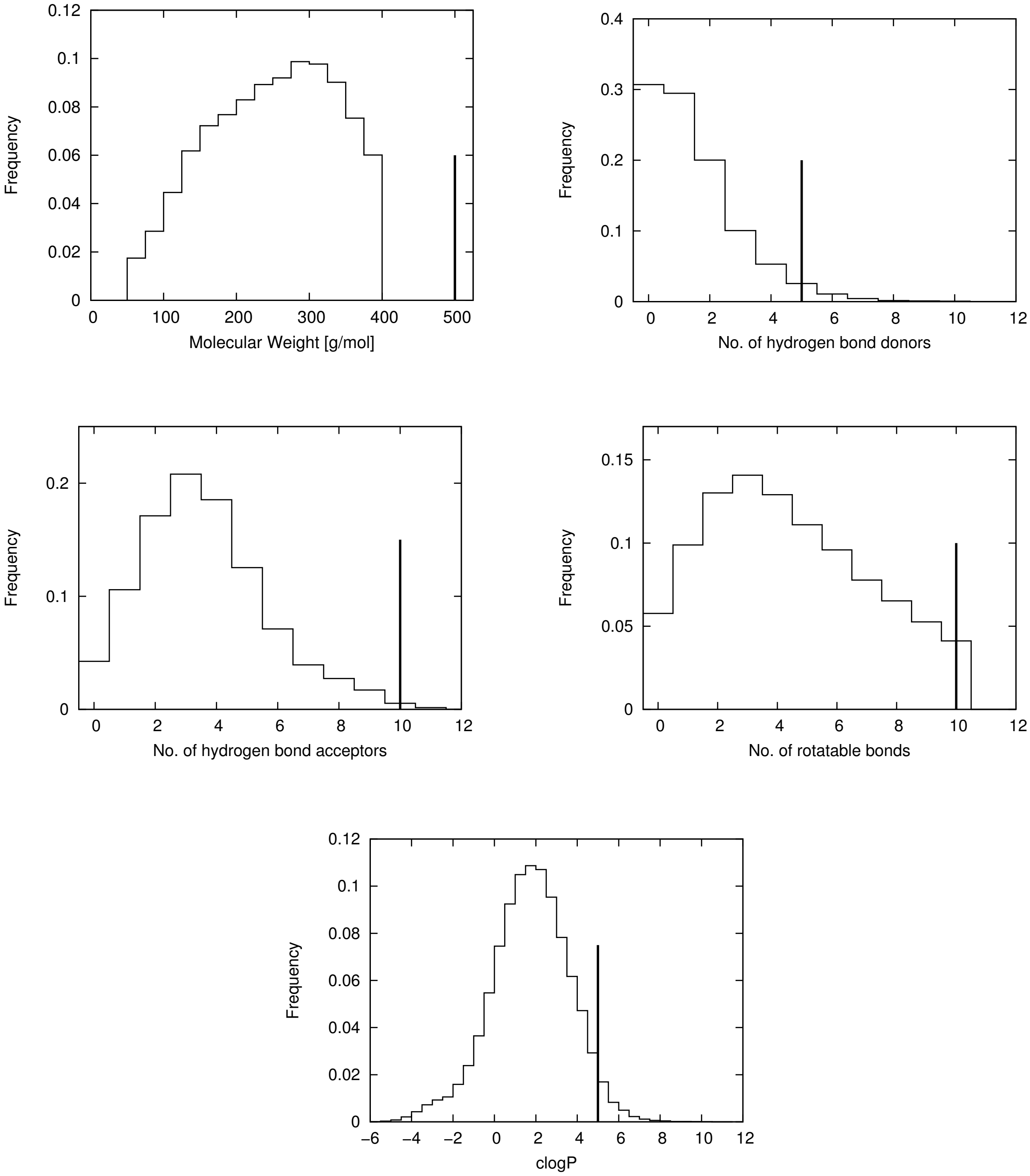


Figure 2.2: Normalized distributions of physicochemical properties of the NP-derived fragment library used for virtual screening. The vertical black lines represent the thresholds defined by Lipinski's rule of five. [42]

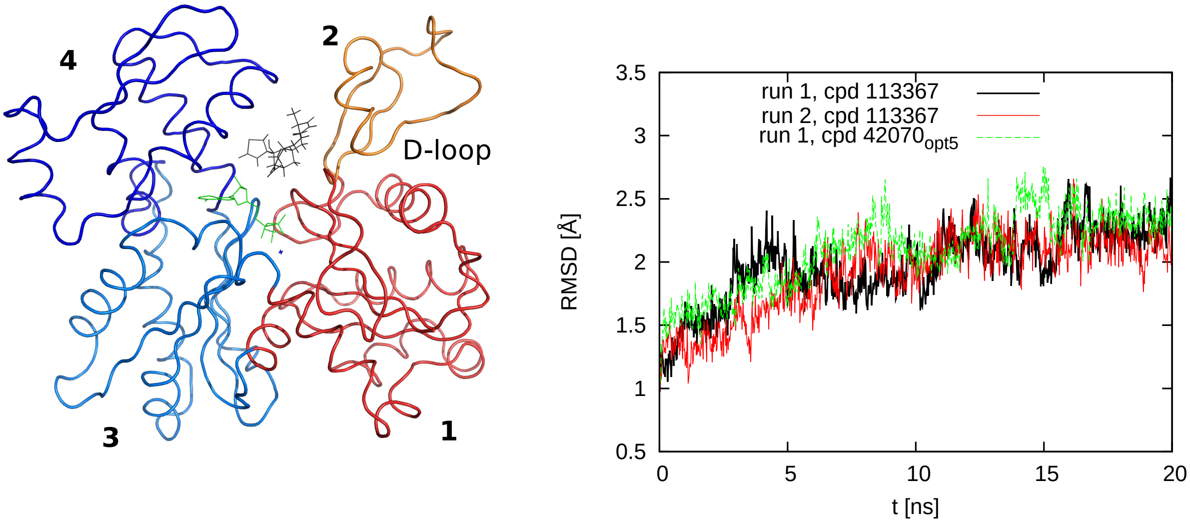


Figure 2.3: (Left) Structure of G-actin in complex with latrunculin A based on the 1ESV crystal structure. Actin is conventionally divided into two major domains. Domain 1 consists of subdomains 1 (shown in red) and 2 (orange), while domain 2 consists of subdomains 3 (light blue) and 4 (dark blue). ATP (green) and latrunculin (gray) bind in a cleft between the two domains. (Right) Time series of RMSD from the actin crystal structure for two runs of actin in complex with compound 113367 (black and red line) and one run for the complex with compound 42070_{opt5} (green line). All C_α atoms (except the D-loop) were used for alignment and calculation of the RMSD. Similar RMSD time series are obtained for other ligands.

Ligand	No. of runs[a]	No. of unbinding events[b]	Unbinding time (leave-one-out)[c] [ns]	Unbinding time (block average)[d] [ns]	Charge
34985	10	9	7 ± 1	5, 7	0
37644	15	1	< 5 events	< 8 events	0
42070	20	6	47 ± 2	< 8 events	-1
$42070_{methoxy}$	20	13	20 ± 1	14, 24	-1
42070_{opt1}	20	3	< 5 events	< 8 events	-1
42070_{opt1_deriv1}	10	1	< 5 events	< 8 events	-1
42070_{opt1_deriv2}	10	4	< 5 events	< 8 events	-1
42070_{opt2}	20	7	48 ± 2	< 8 events	-1
42070_{opt2_5ring}	10	1	< 5 events	< 8 events	-1
42070_{opt3}	20	7	48 ± 2	< 8 events	-1
42070_{opt5}	20	6	51 ± 1	< 8 events	-1
42070_{phe}	10	4	< 5 events	< 8 events	-1
42070_{tyr}	10	4	< 5 events	< 8 events	-1
42430	20	5	76 ± 5	< 8 events	-1
42783	20	0	< 8 events	< 8 events	-2
102213	16	7	36 ± 1	< 8 events	-1
$102213_{enantio}$	20	7	46 ± 1	< 8 events	-1
104690	20	14	23 ± 1	18, 25	0
105302	15	0	< 5 events	< 8 events	-2
105339	20	12	25 ± 1	23, 24	0
106230	20	6	66 ± 4	< 8 events	0
106499	15	3	< 5 events	< 8 events	-1
106499_{deriv1}	20	9	31 ± 1	21,38	-1
$106499_{deriv2\_enantio1}$	10	1	< 5 events	< 8 events	-1
$106499_{deriv2\_enantio2}$	10	0	< 5 events	< 8 events	-1
106540	20	6	56 ± 3	< 8 events	0
108360	20	9	42 ± 1	28,49	0
109412	20	7	46 ± 2	< 8 events	0
113367	20	6	76 ± 4	< 8 events	0

Table 2.1: [a]Multiple runs were started using different seeds to generate a random distribution of the initial velocities. [b]An unbinding event is defined as a separation of the ligand center of mass from the center of mass of the binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210) larger than 10 Å. [c]Leave-one-out unbinding times were calculated for molecules with at least five dissociation events by performing x single exponential fits, where x is the number of unbinding events. The average values of unbinding time and rms error are reported in the table. [d]Block averaging was performed for molecules with at least eight unbinding events by partitioning the 10-20 individual runs of one compound into two blocks of equal size. The arithmetic mean of the unbinding time calculated for each of the two blocks is given in the table.

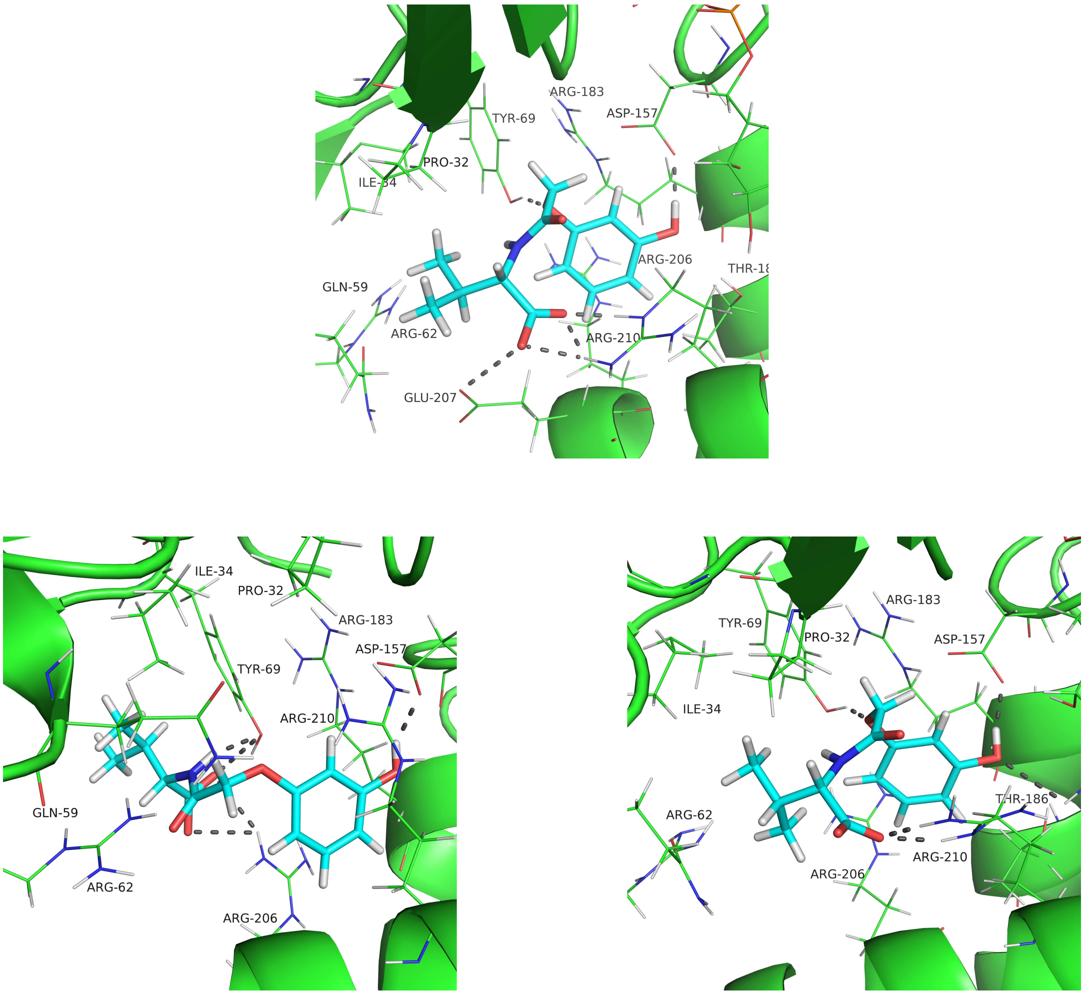


Figure 2.4: (Top) Binding mode of compound 42070 predicted by docking. (Bottom) Binding mode of compound 42070 according to the most populated (left) and second most populated (right) cluster in MD simulations.

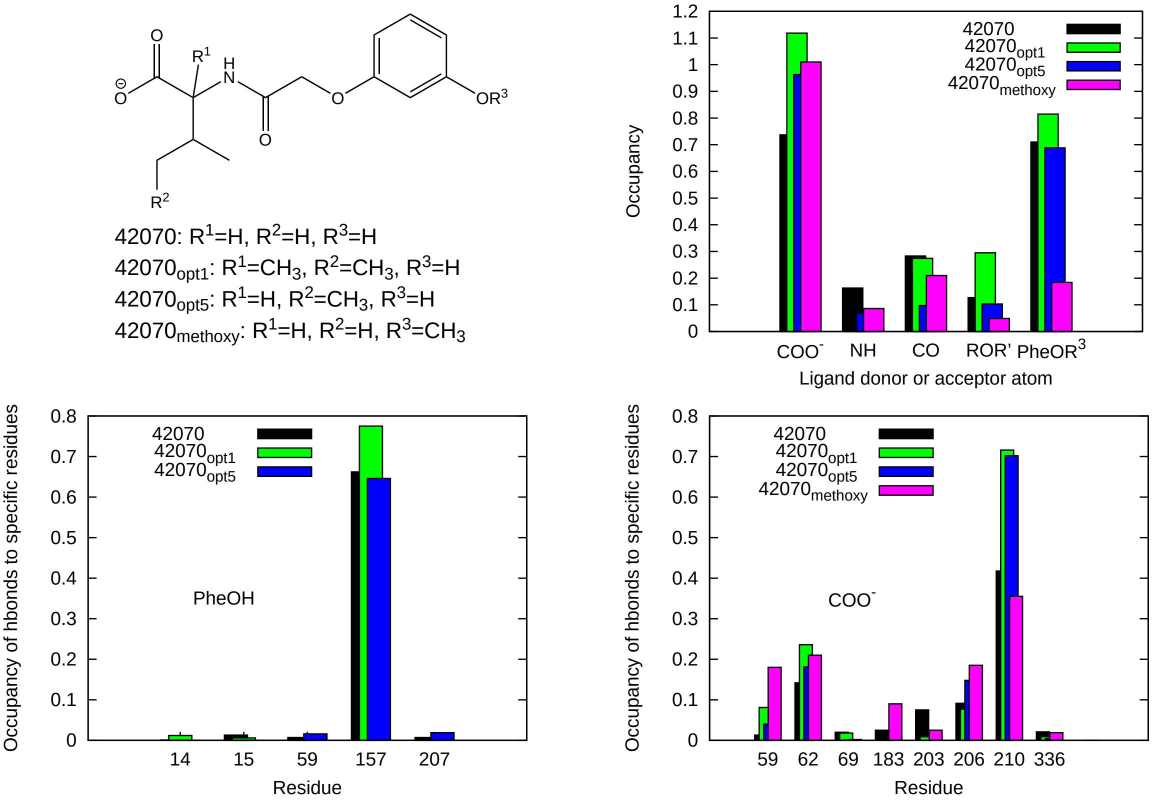


Figure 2.5: (Top left) Chemical structure of compounds 42070, 42070_{opt1}, 42070_{opt5}, and $42070_{methoxy}$. (Top right) Occupancy of hydrogen bond donors and acceptors, i.e., the percentage of simulation time where a hydrogen bond is formed, for compounds 42070, 42070_{opt1}, 42070_{opt5}, and $42070_{methoxy}$ averaged over all runs of each compound. The carboxylate group is labeled as COO^-, the amide nitrogen and oxygen as NH and CO, respectively, the ether group as ROR', and the phenolic hydroxyl or methoxy group as $PheOR^3$. (Bottom left) Average occupancy of different residues involved in a hydrogen bond to the phenolic hydroxyl group of the ligands. (Bottom right) Average occupancy of different residues involved in a hydrogen bond to the carboxylate group of the ligands.

Supplementary Material for:
Targeting G-actin by virtual screening of a natural product-inspired library and hit validation by molecular dynamics simulations

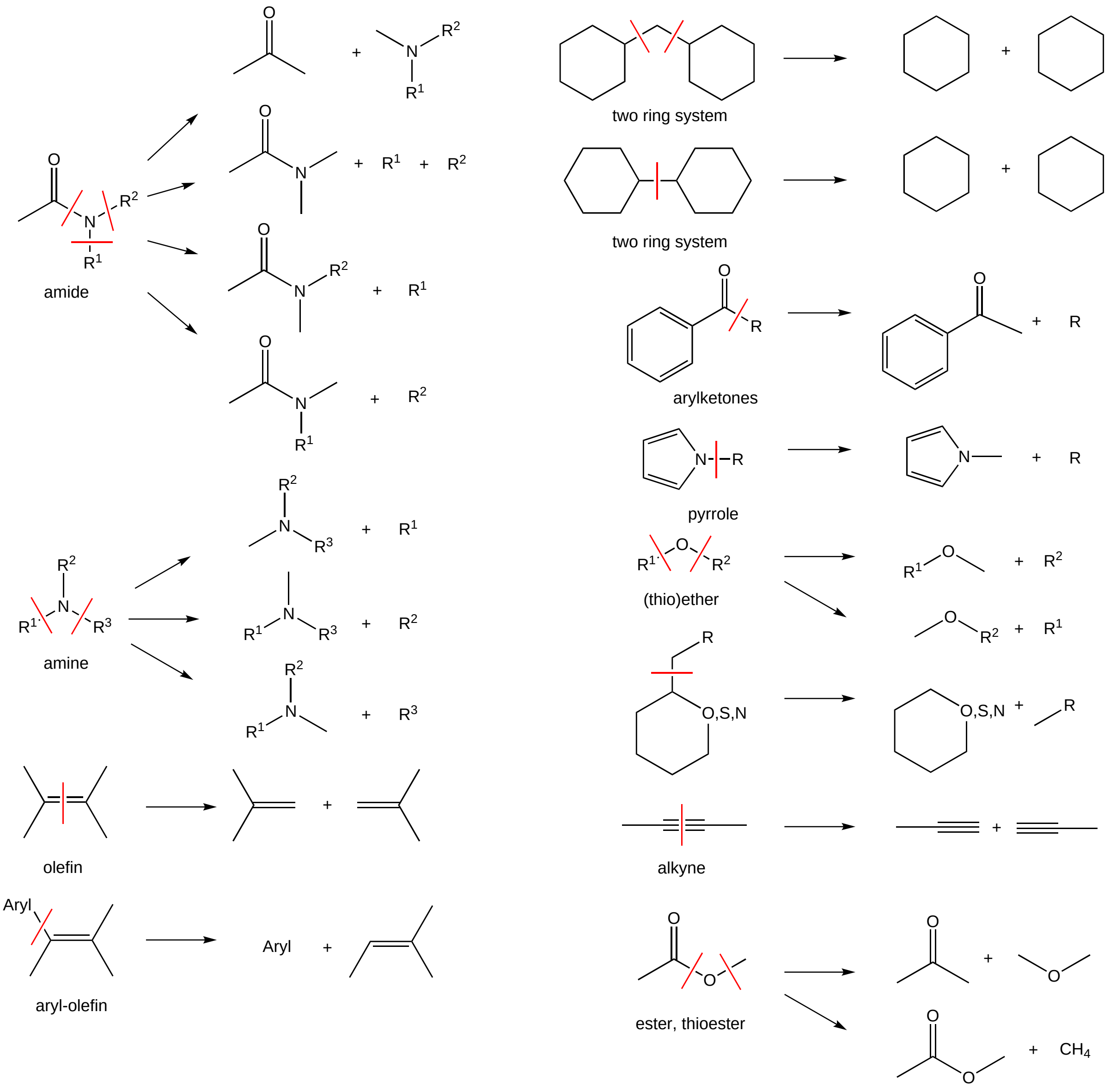
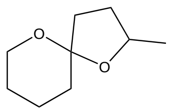
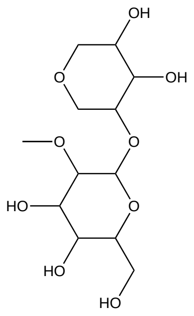
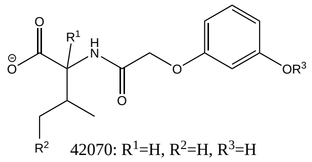
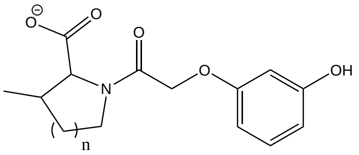


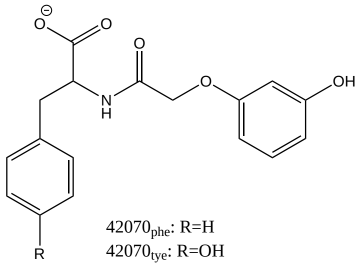
Figure S1: Rules for decomposition of compounds according to organic synthesis. A methyl group is used to fill valences of electronegative atoms to avoid the creation of unwanted hydrogen bond donors. The hybridization states of sp and sp^2 hybridized atoms is retained to not change the electronic environment.

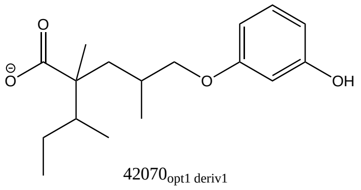

34985

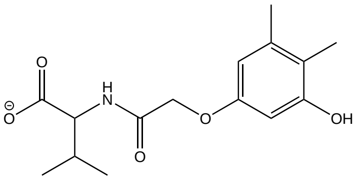

37644

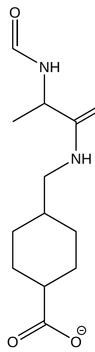

42070: R^1=H, R^2=H, R^3=H
42070_{methoxy}: R^1=H, R^2=H, R^3=CH_3
42070_{opt1}: R^1=CH_3, R^2=CH_3, R^3=H
$42070_{\text{opt1 deriv2}}$: R^1=CH_2CH_3, R^2=CH_3, R^3=H
42070_{opt5}: R^1=H, R^2=CH_3, R^3=H

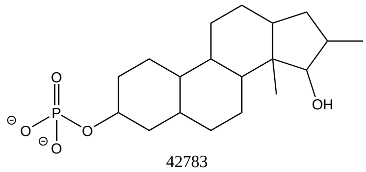

$42070_{\text{opt2 5ring}}$: n=1
42070_{opt2}: n=2

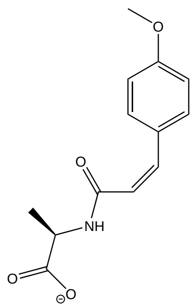

42070_{phe}: R=H
42070_{tye}: R=OH


$42070_{\text{opt1 deriv1}}$


42070_{opt3}


42430


42783


102213

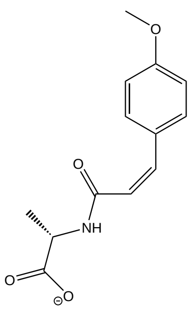
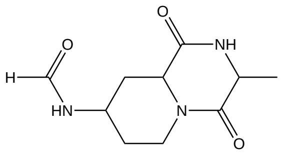
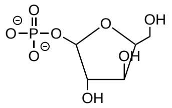
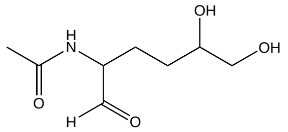
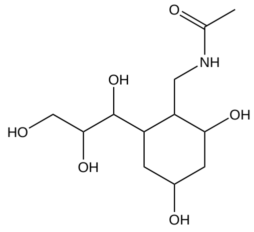

102213_{enantio}

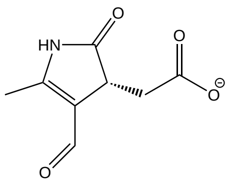
Figure S2: Chemical structures of compounds 1-17 selected for MD analysis

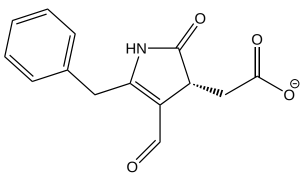

104690

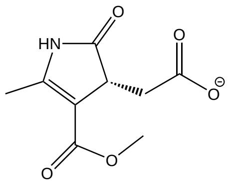

105302

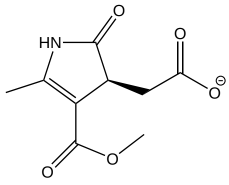

105339

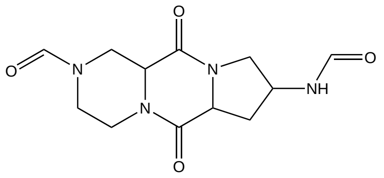

106230

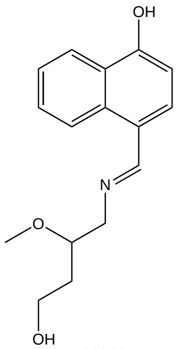

106499

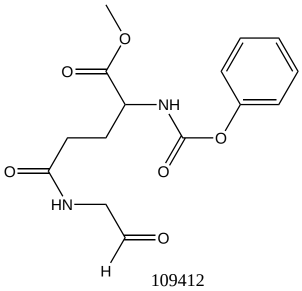

106499_{deriv1}


$106499_{\text{deriv2\_enantio1}}$


$106499_{\text{deriv2\_enantio2}}$


106540


108360


109412

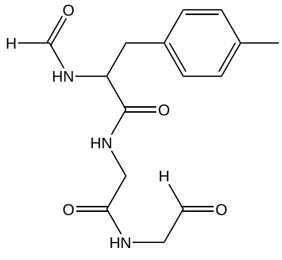

113367

Figure S3: Chemical structures of compounds 18-29 selected for MD analysis

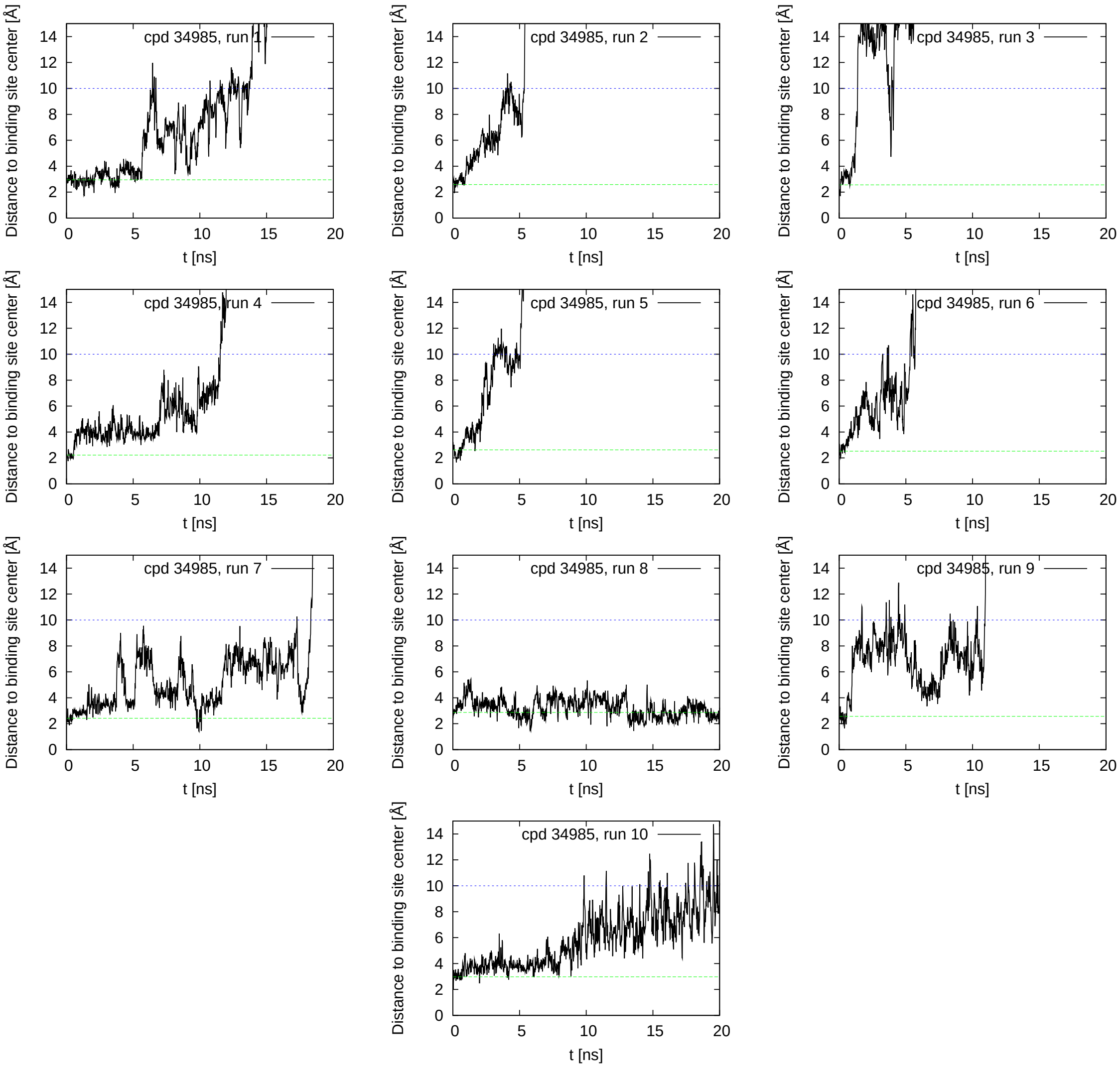


Figure S4: Time series of distance between centers of mass of compound 34985 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

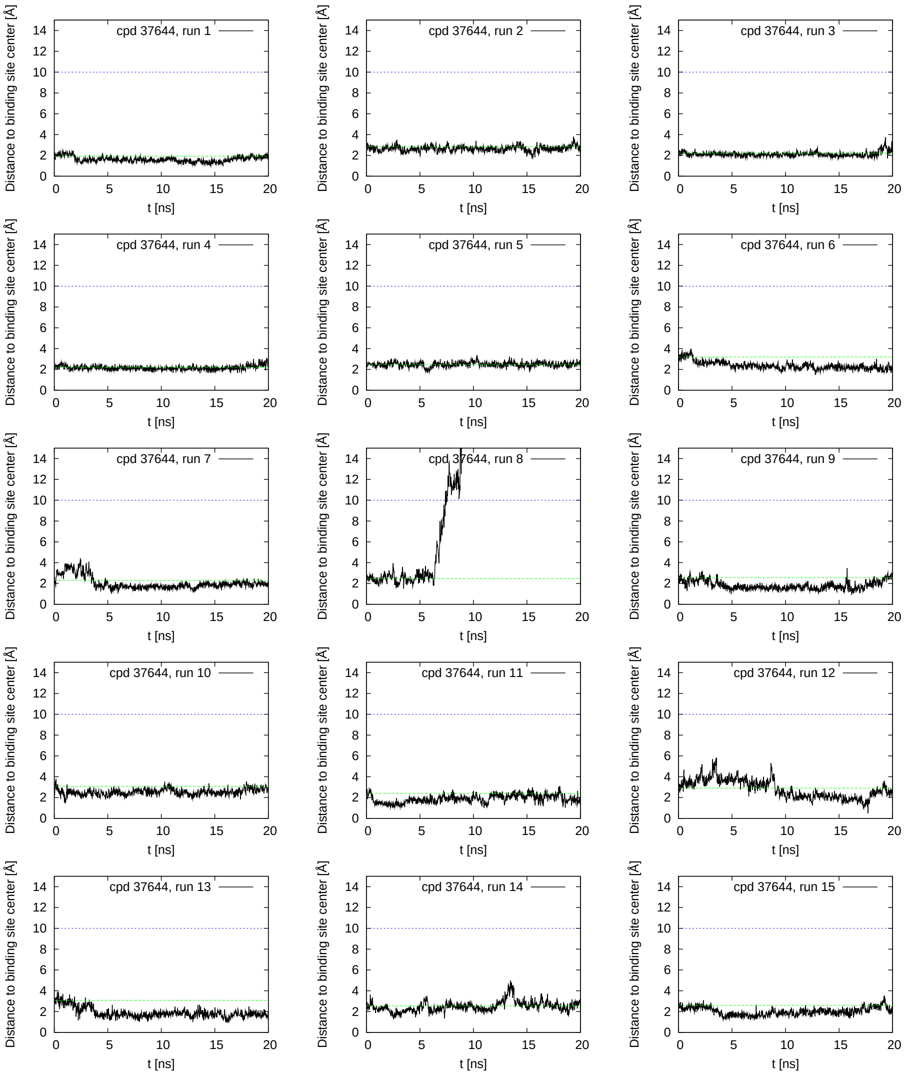


Figure S5: Time series of distance between centers of mass of compound 37644 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

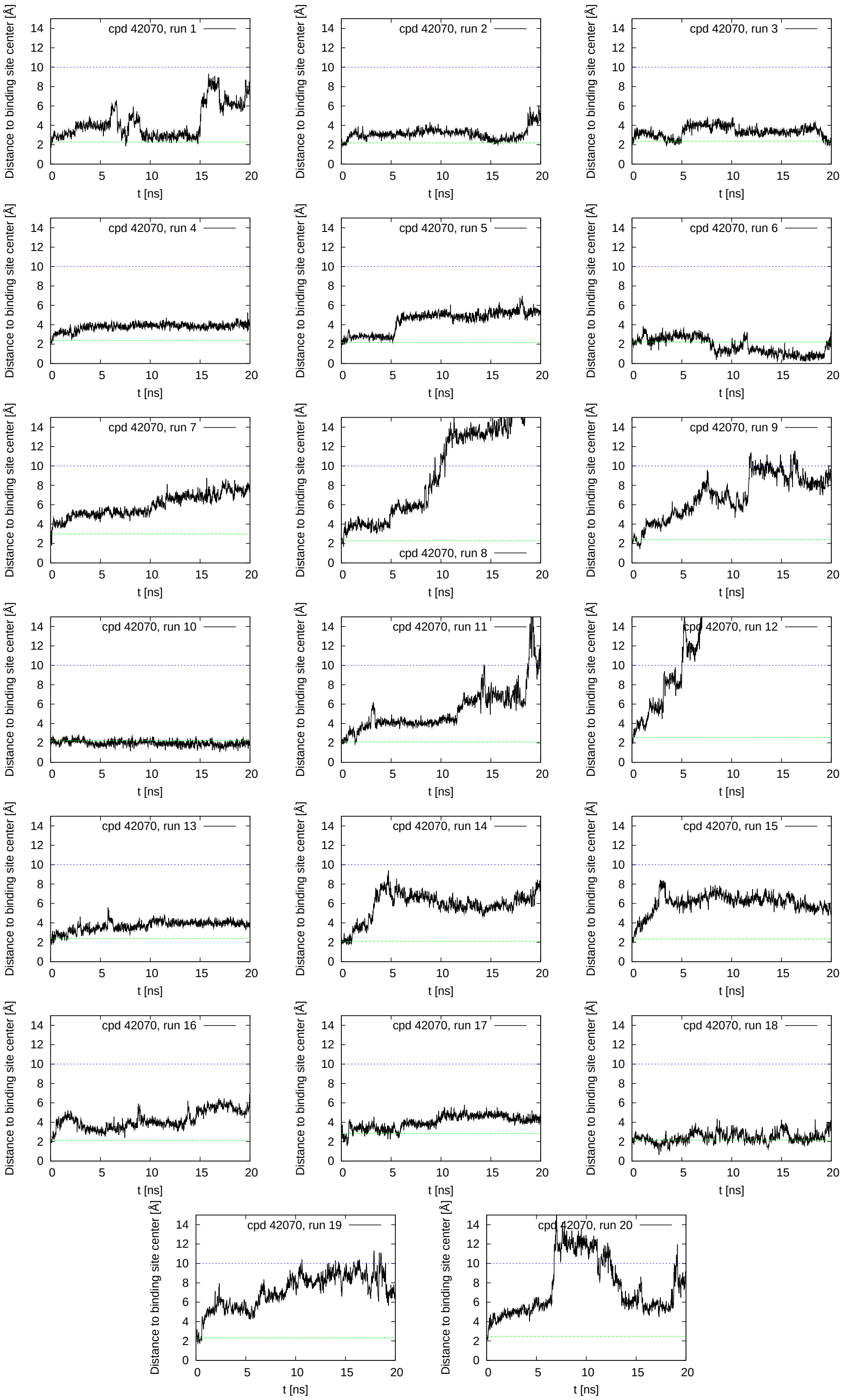


Figure S6: Time series of distance between centers of mass of compound 42070 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

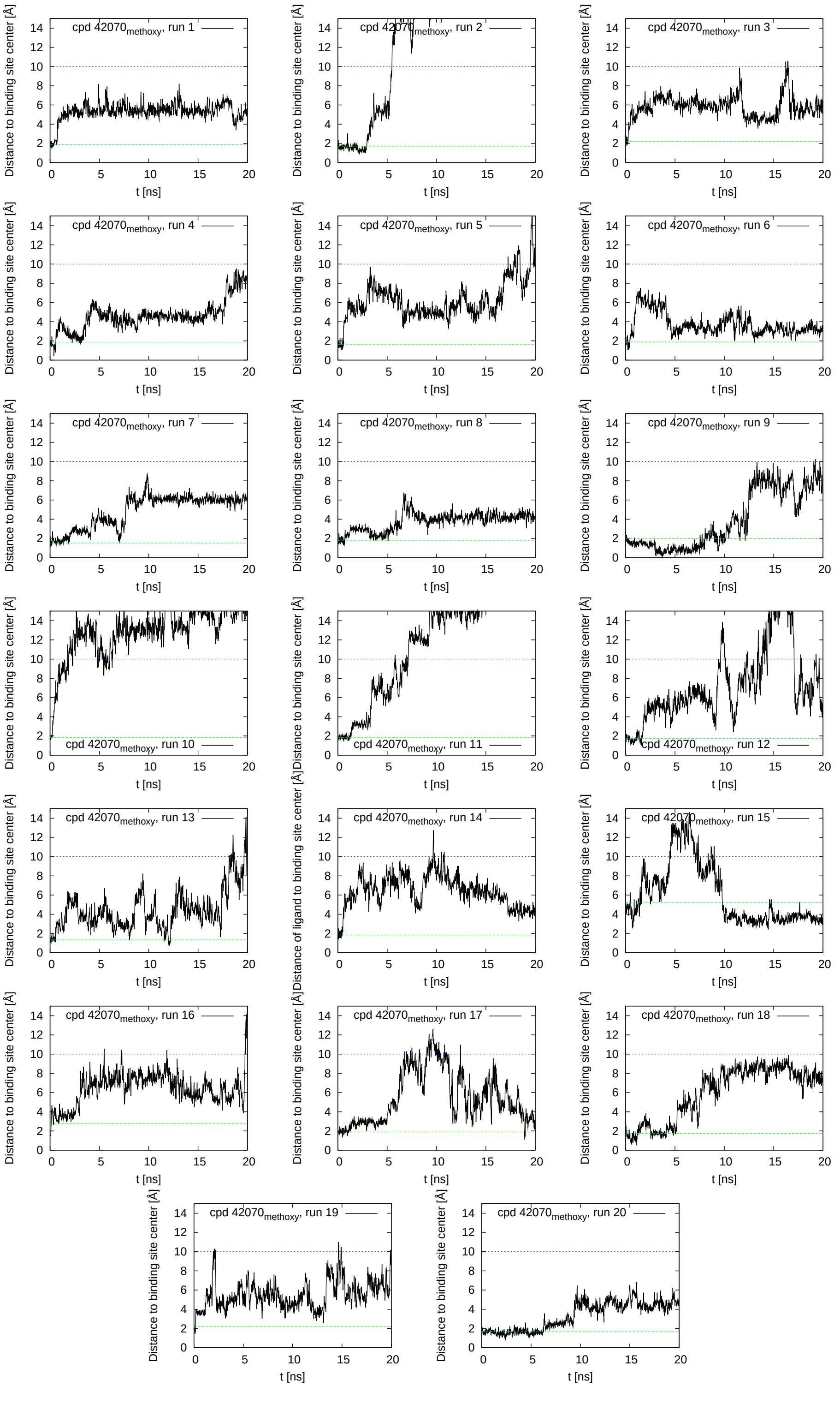


Figure S7: Time series of distance between centers of mass of compound $42070_{methoxy}$ and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

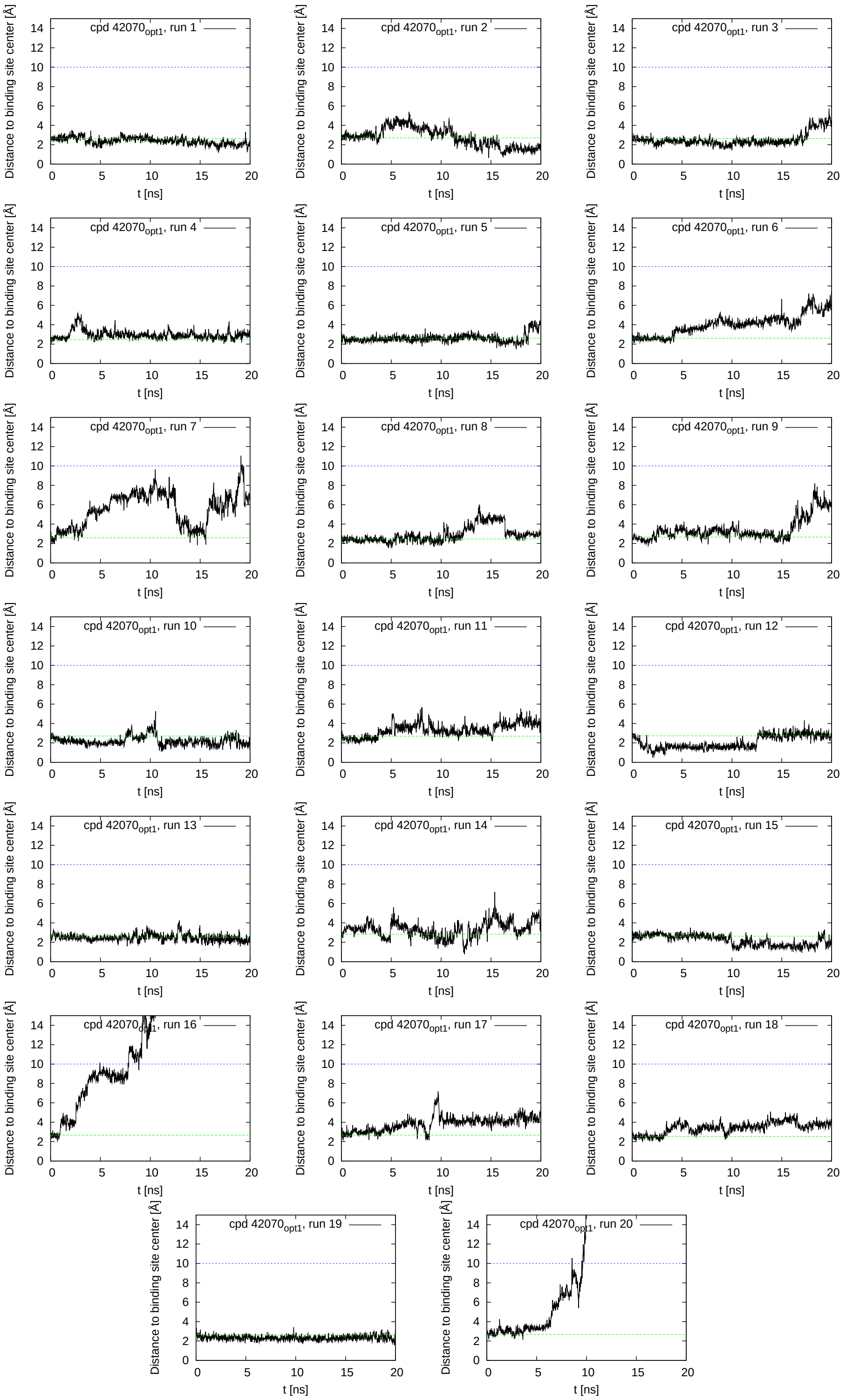


Figure S8: Time series of distance between centers of mass of compound 42070_{opt1} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

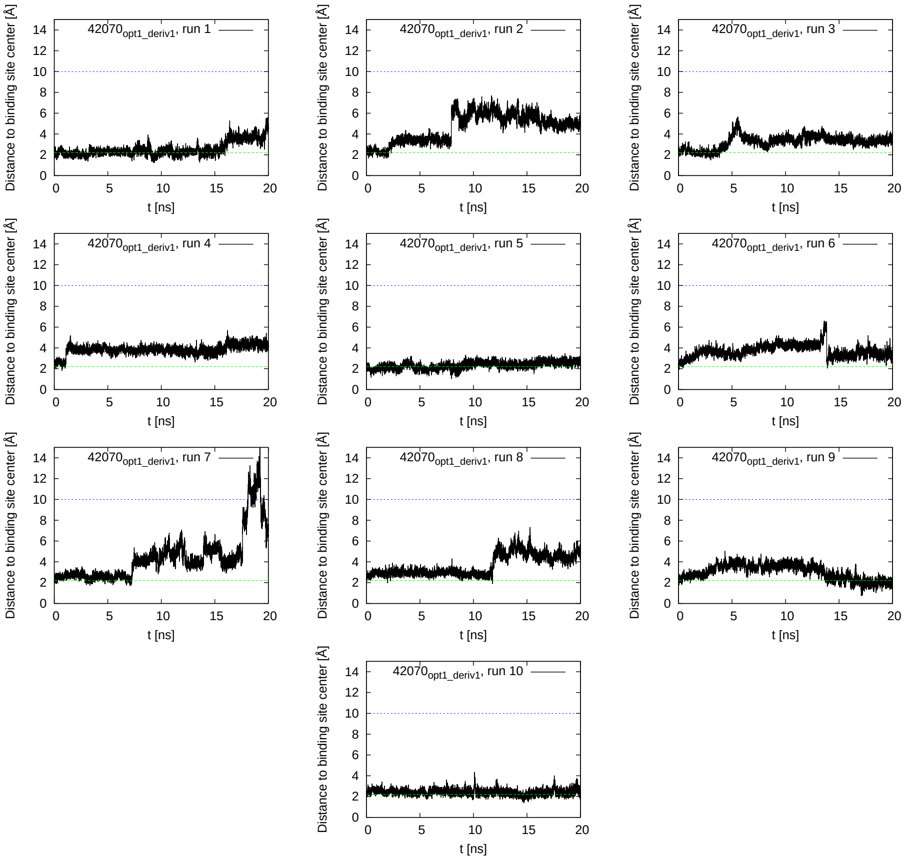


Figure S9: Time series of distance between centers of mass of compound 42070_{opt1_deriv1} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

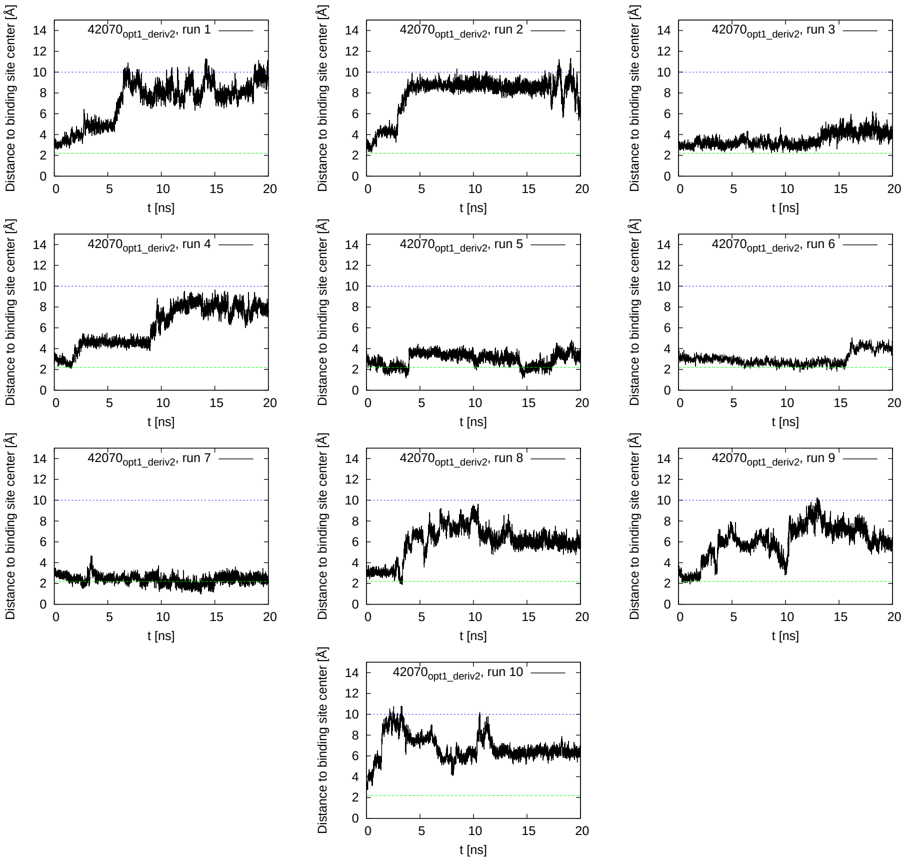


Figure S10: Time series of distance between centers of mass of compound 42070$_{opt1\_deriv2}$ and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

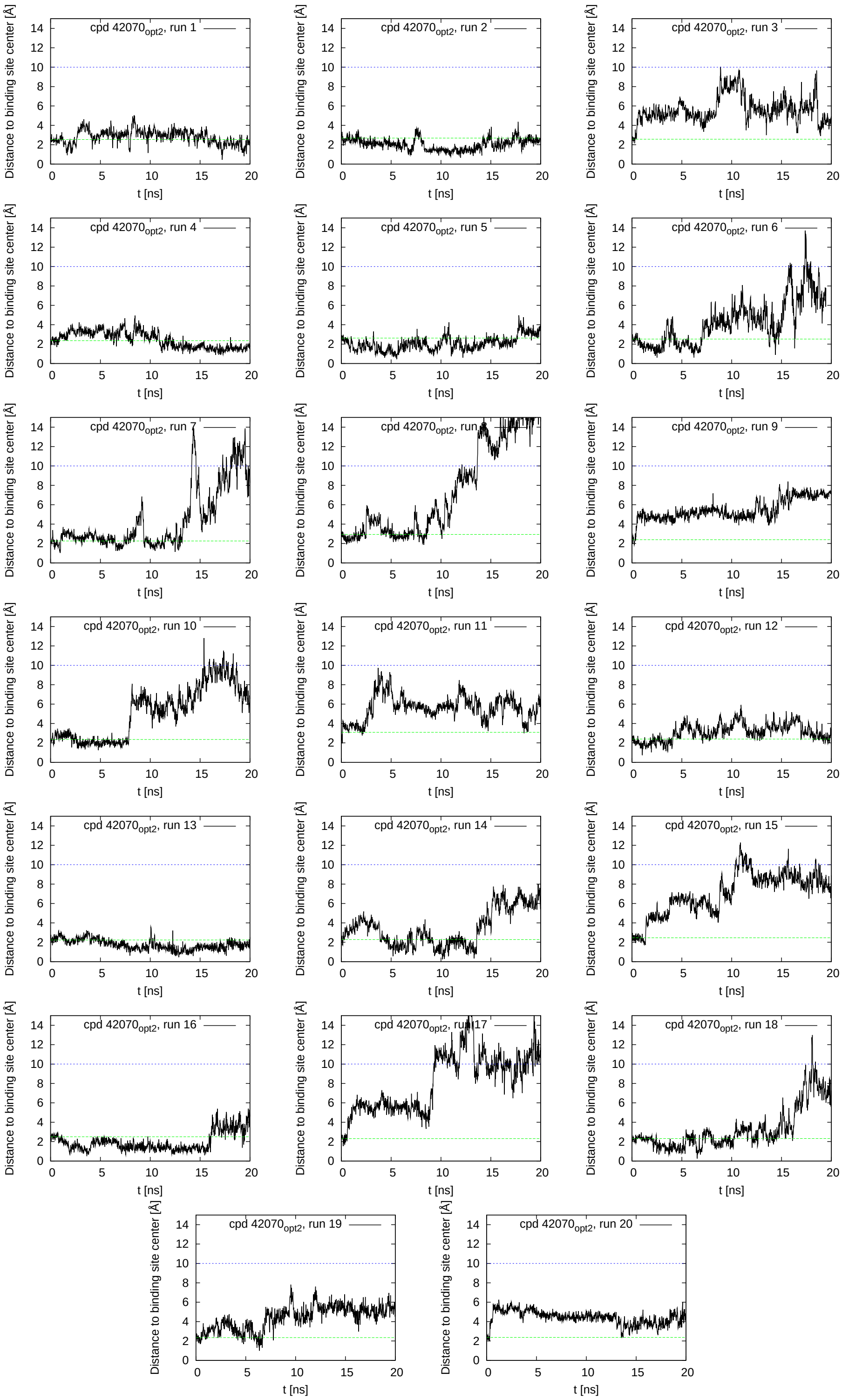


Figure S11: Time series of distance between centers of mass of compound 42070_{opt2} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

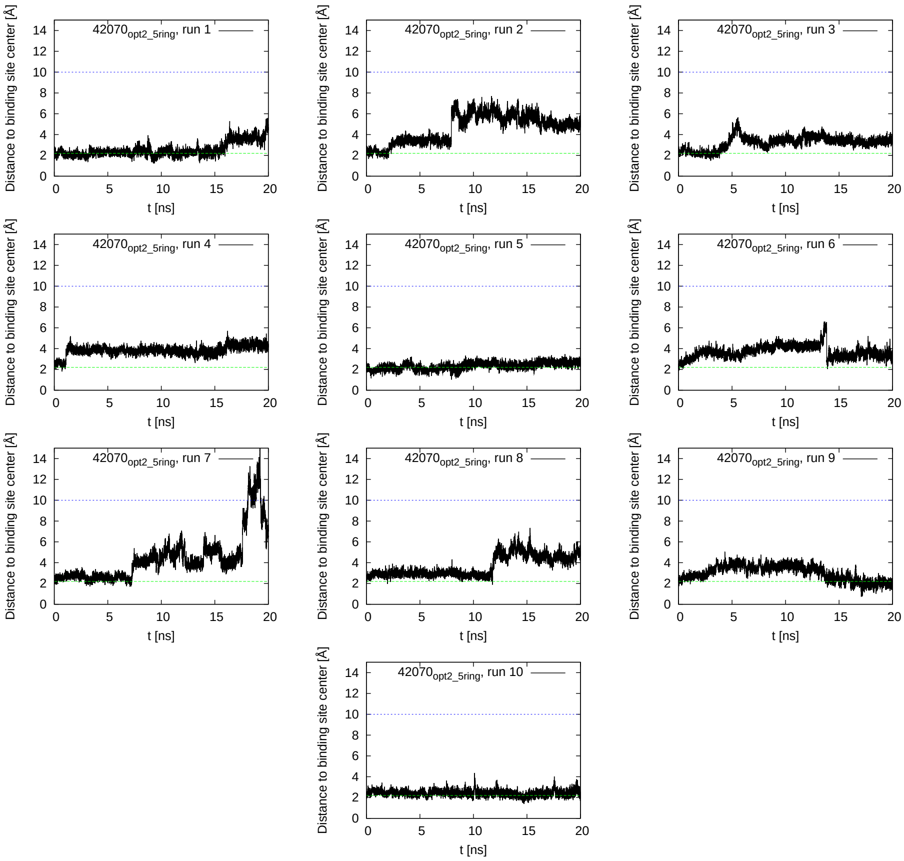


Figure S12: Time series of distance between centers of mass of compound 42070_{opt2_5ring} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

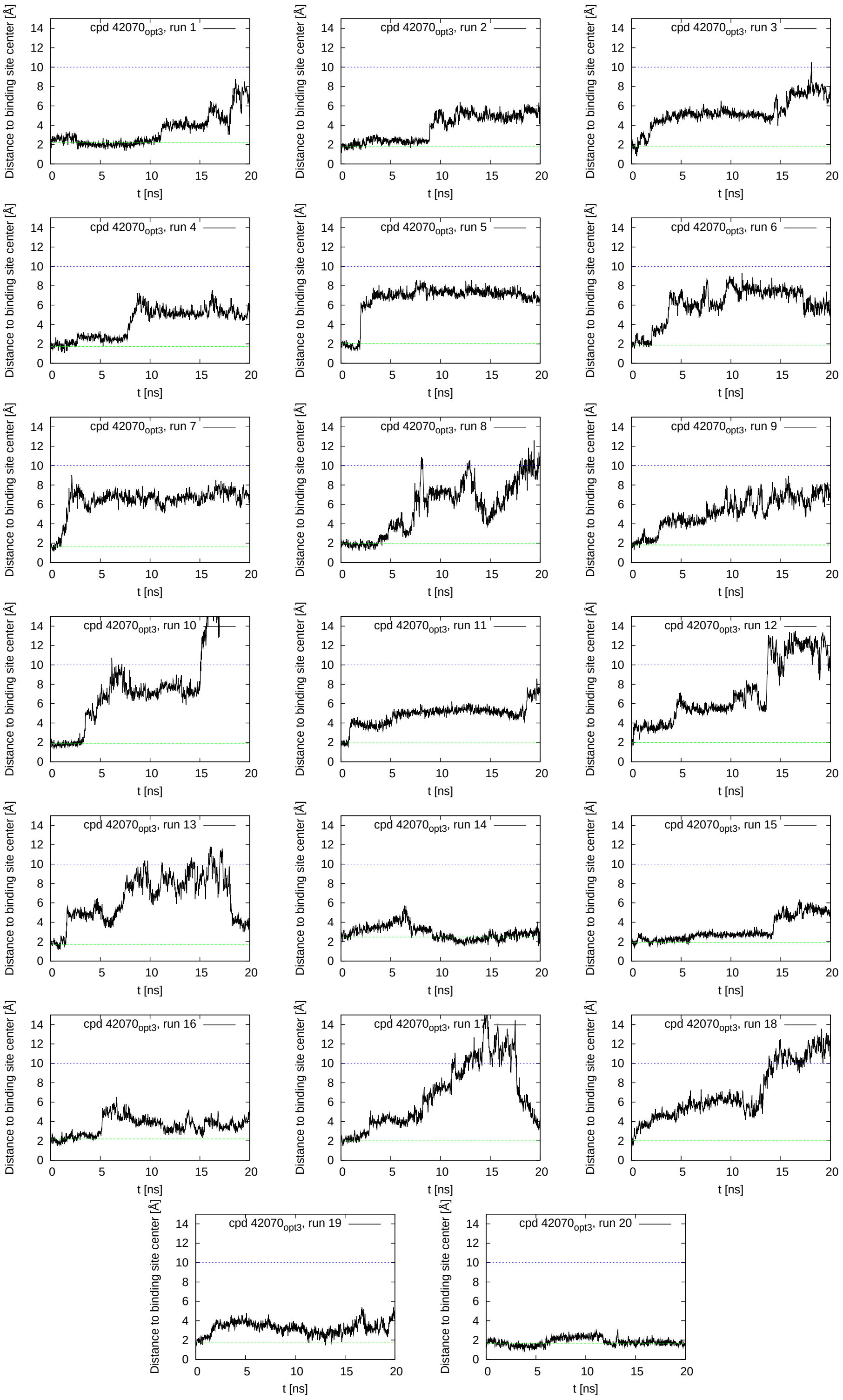


Figure S13: Time series of distance between centers of mass of compound 42070_{opt3} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

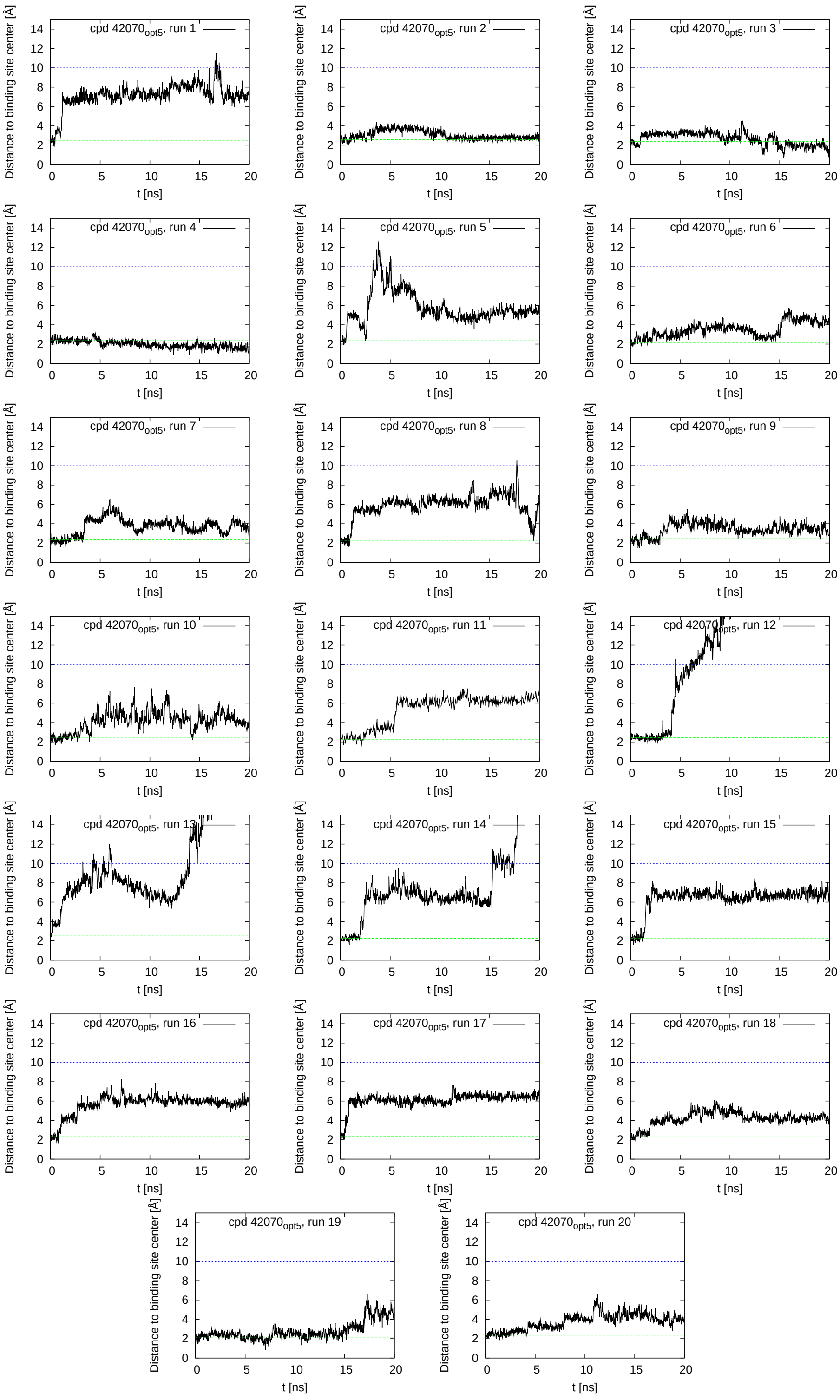


Figure S14: Time series of distance between centers of mass of compound 42070_{opt5} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

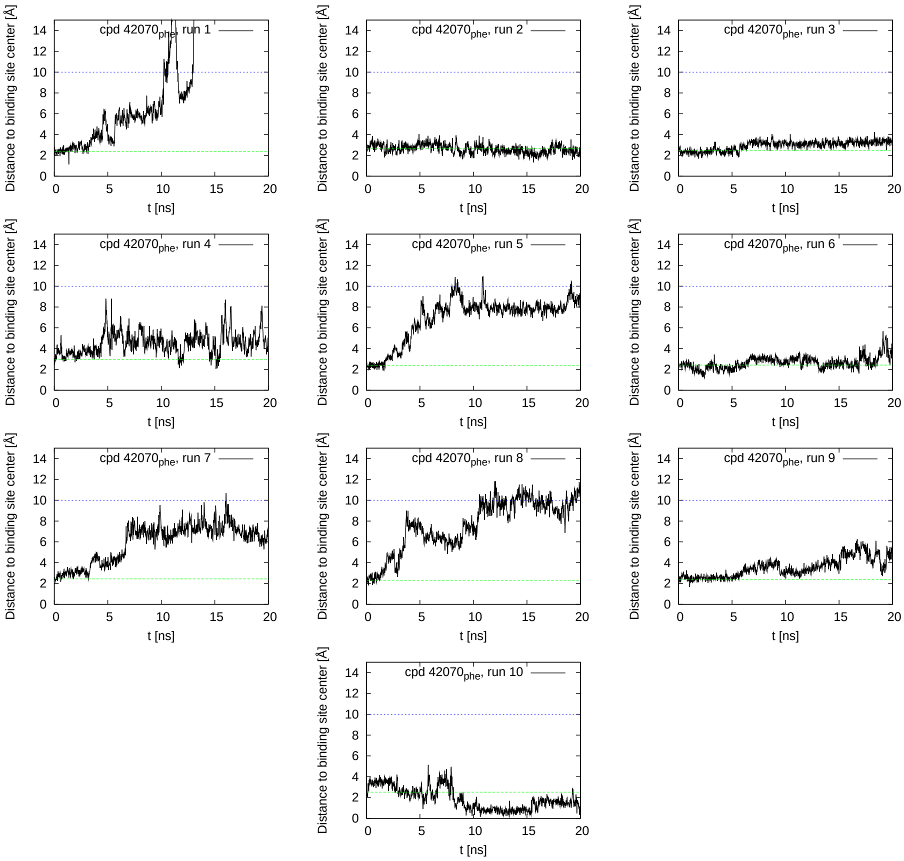


Figure S15: Time series of distance between centers of mass of compound 42070_{phe} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

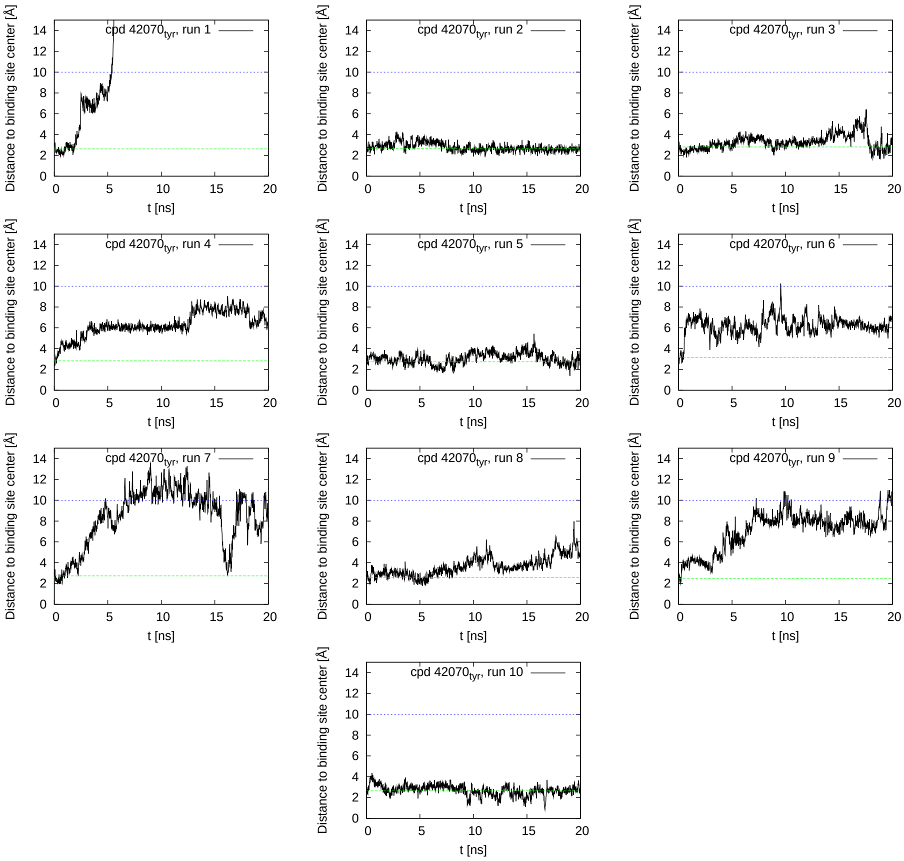


Figure S16: Time series of distance between centers of mass of compound 42070_{tyr} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

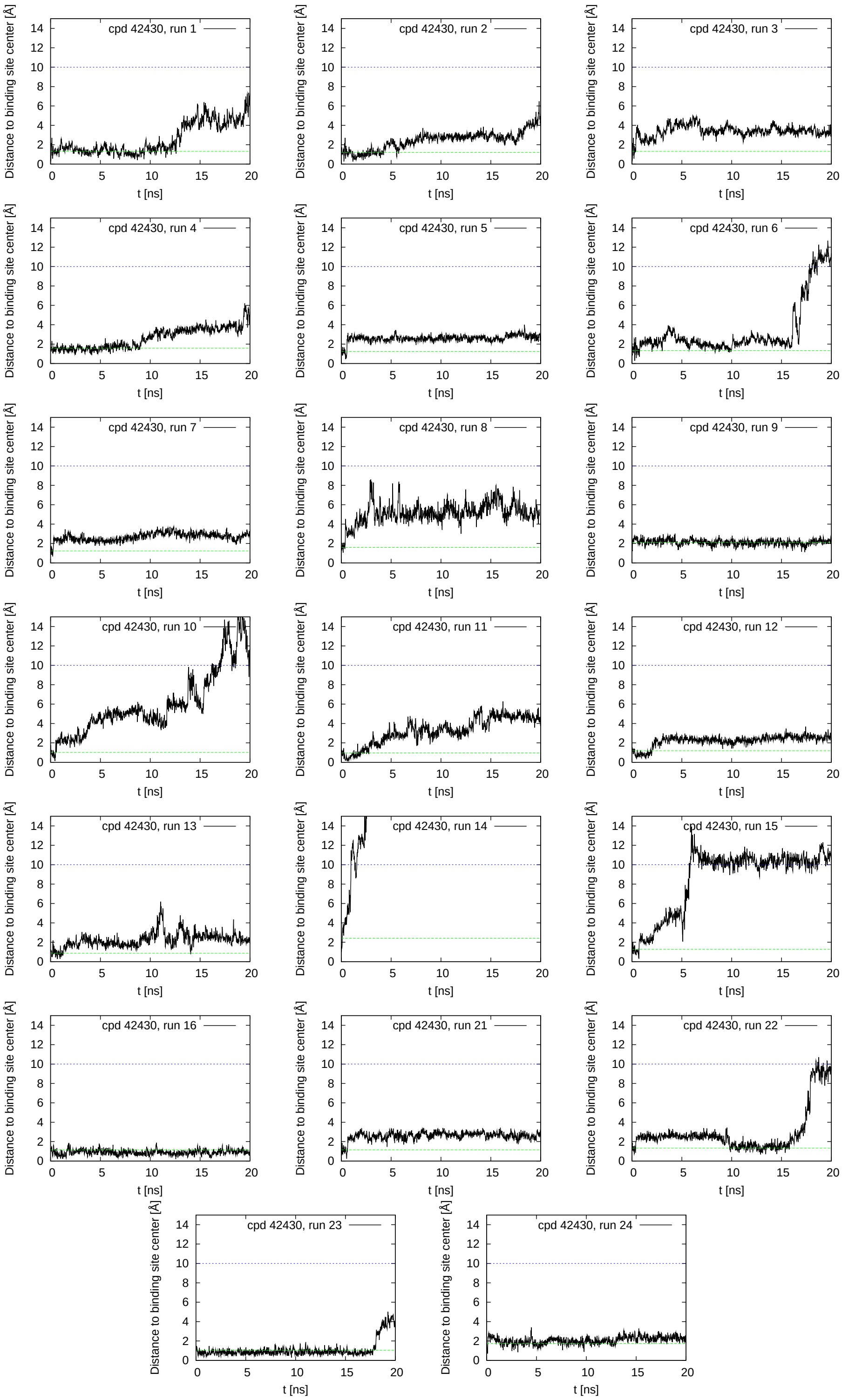


Figure S17: Time series of distance between centers of mass of compound 42430 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

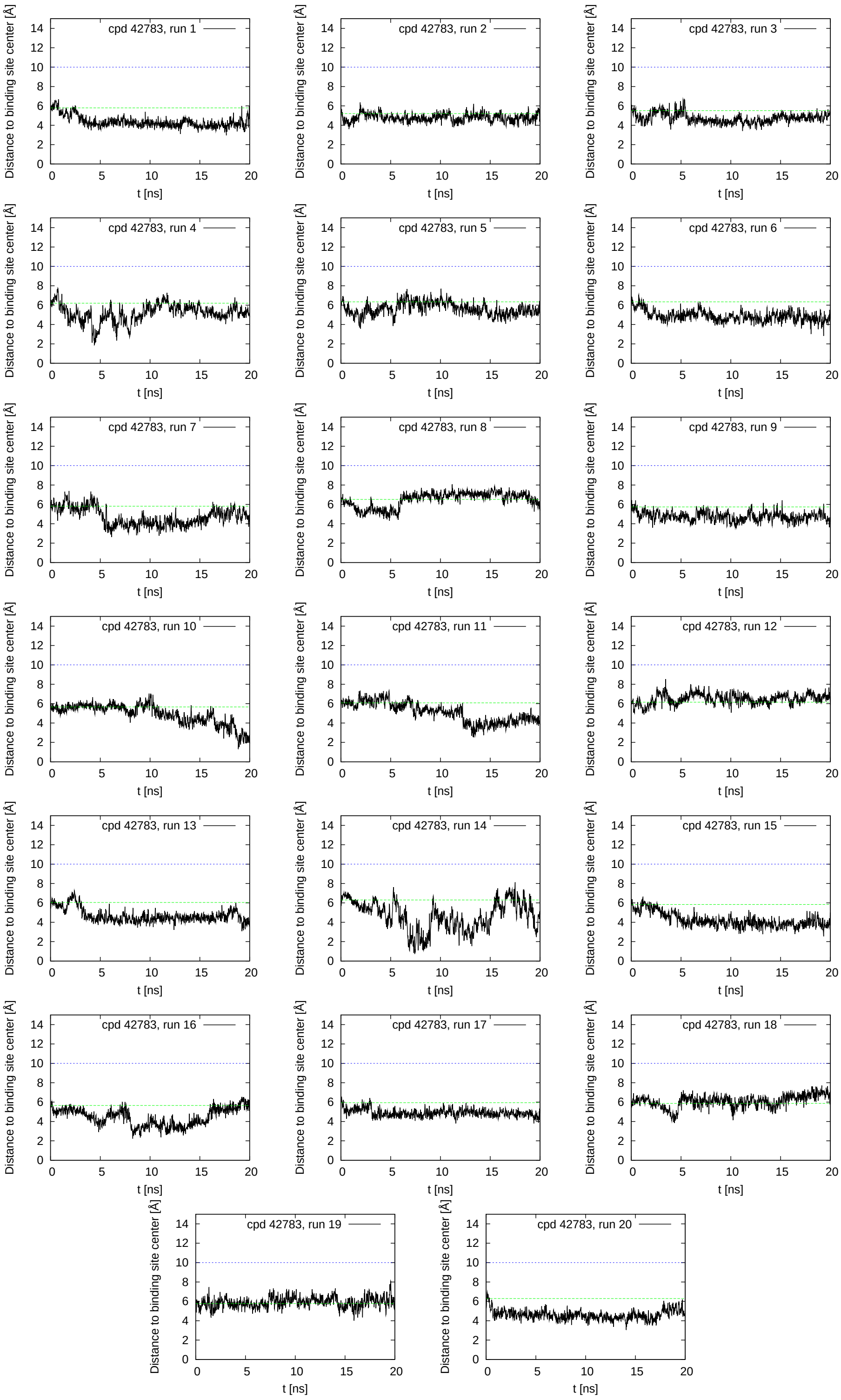


Figure S18: Time series of distance between centers of mass of compound 42783 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

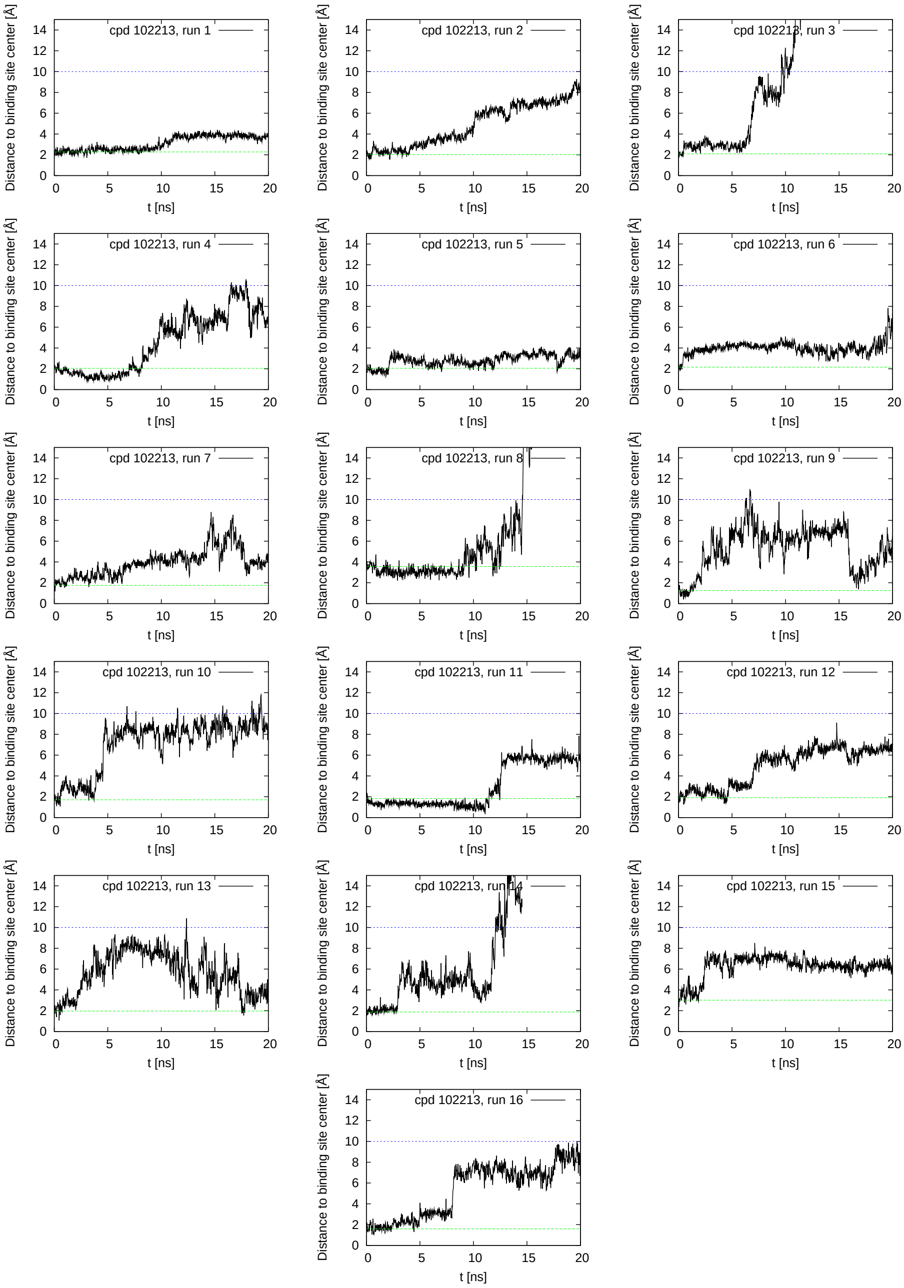


Figure S19: Time series of distance between centers of mass of compound 102213 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

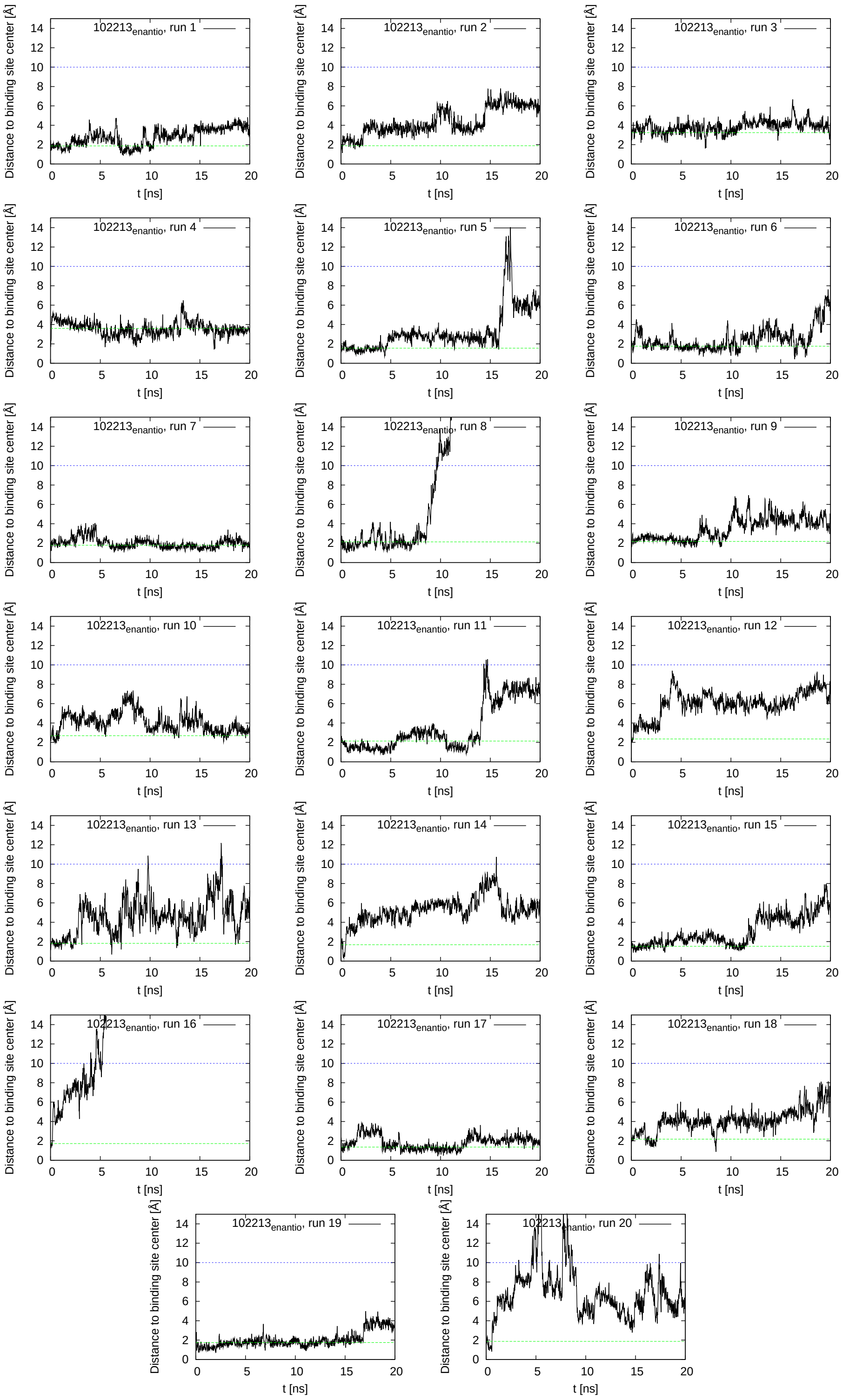


Figure S20: Time series of distance between centers of mass of compound $102213_{enantio}$ and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

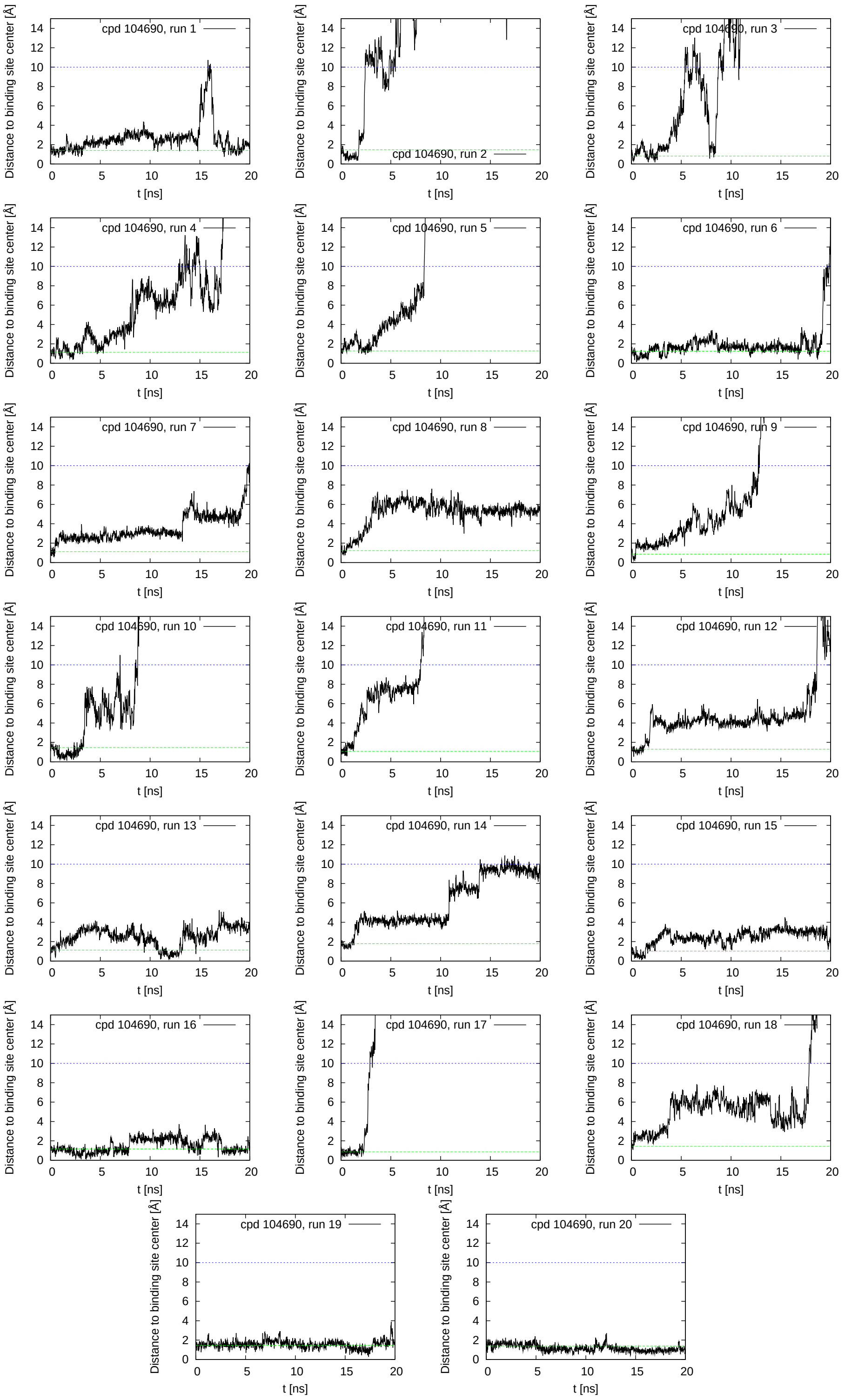


Figure S21: Time series of distance between centers of mass of compound 104690 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

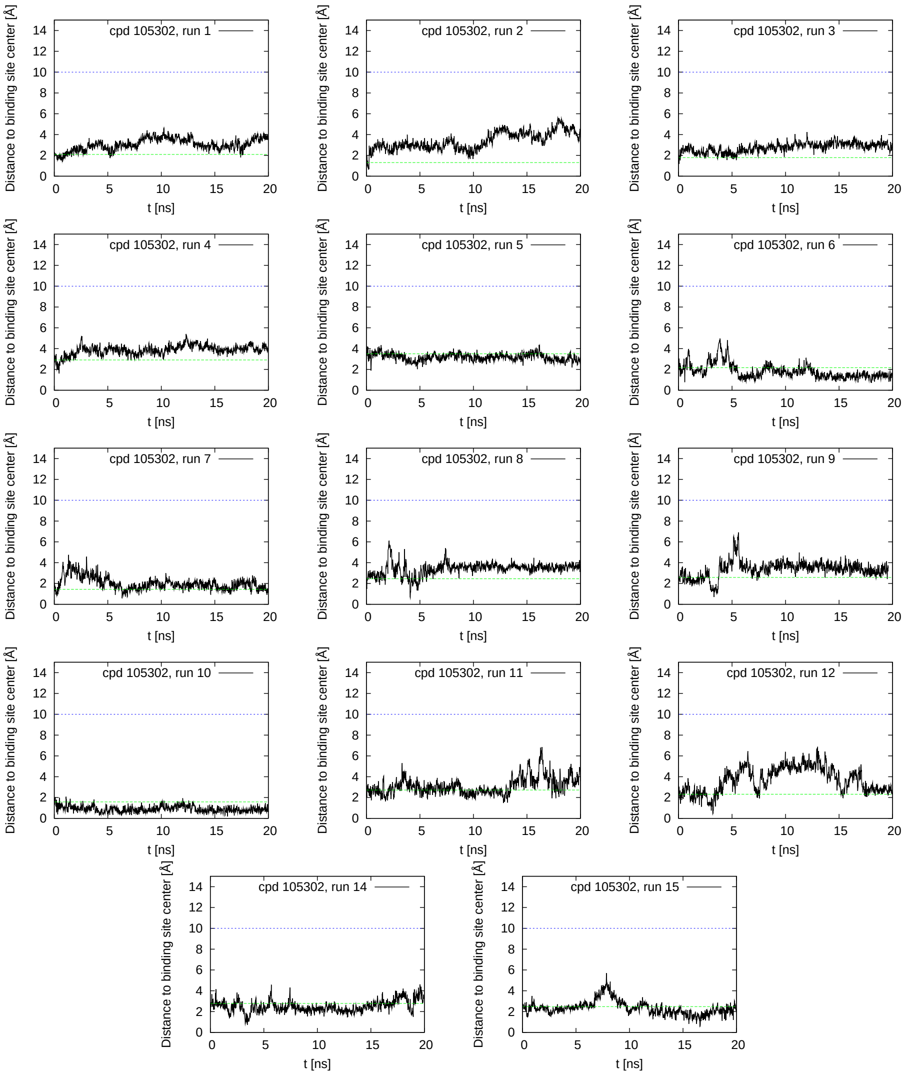


Figure S22: Time series of distance between centers of mass of compound 105302 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

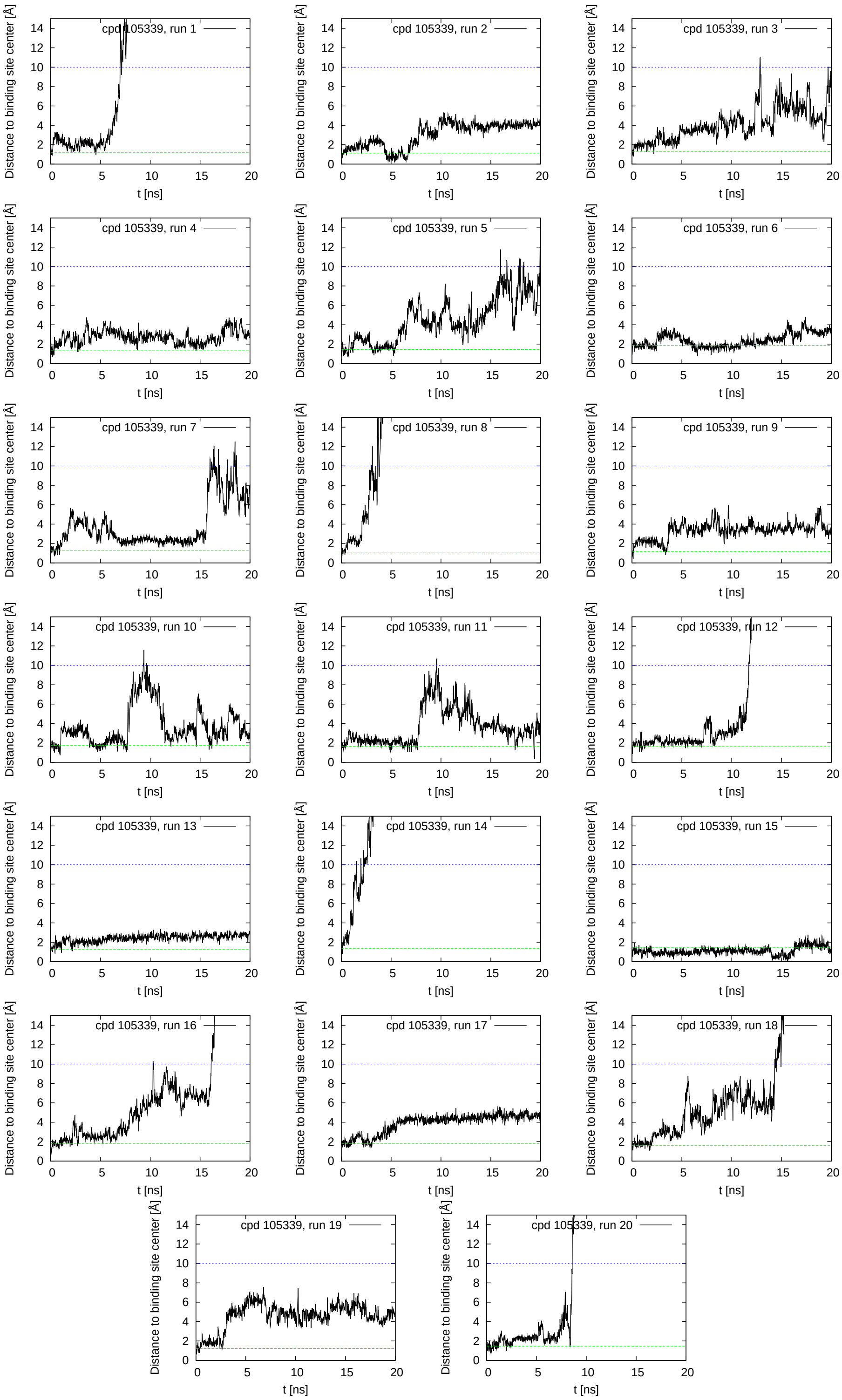


Figure S23: Time series of distance between centers of mass of compound 105339 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

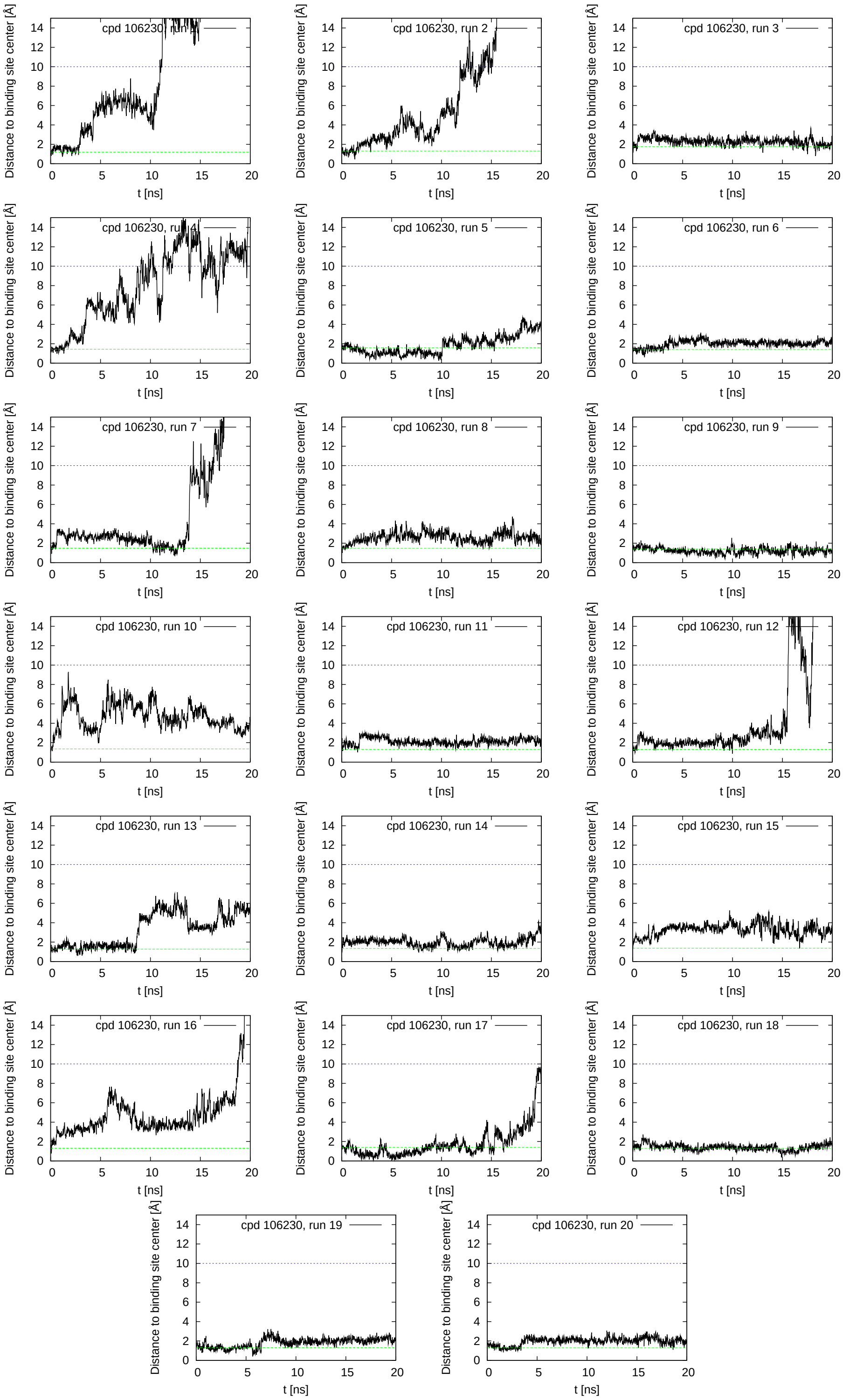


Figure S24: Time series of distance between centers of mass of compound 106230 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

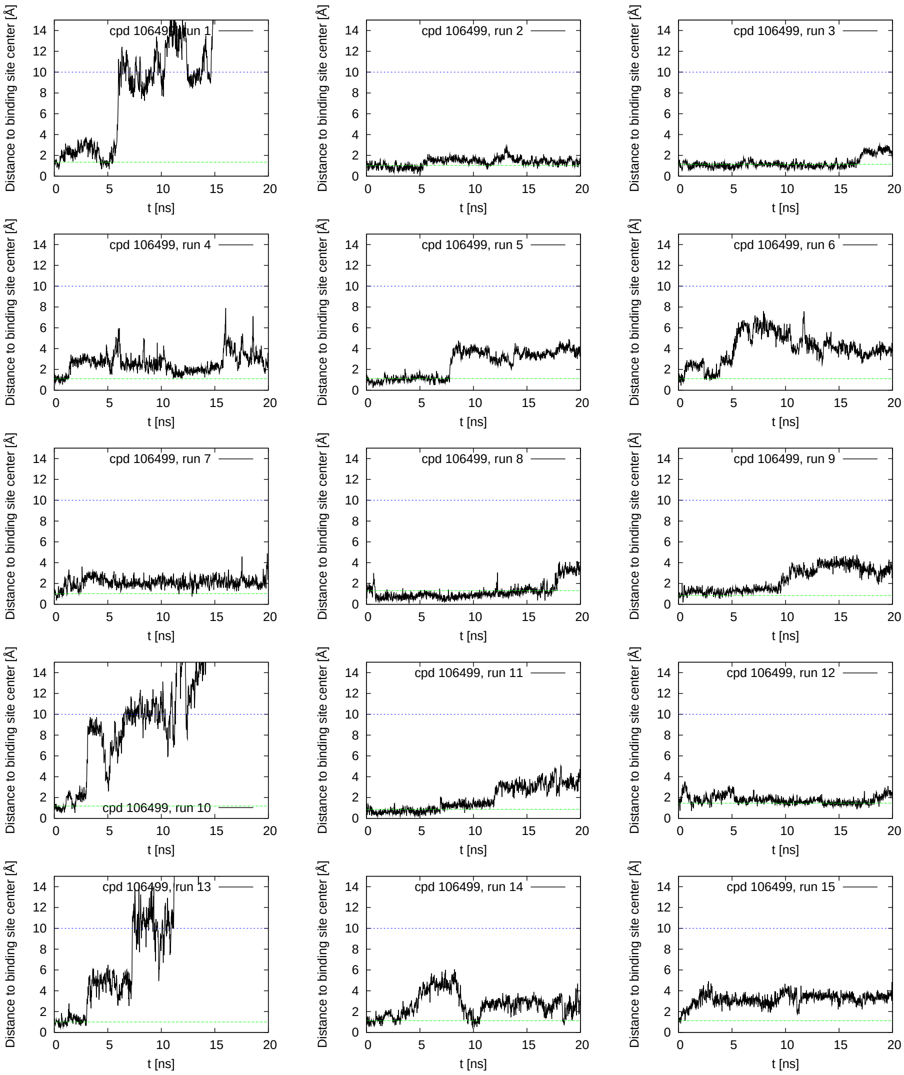


Figure S25: Time series of distance between centers of mass of compound 106499 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

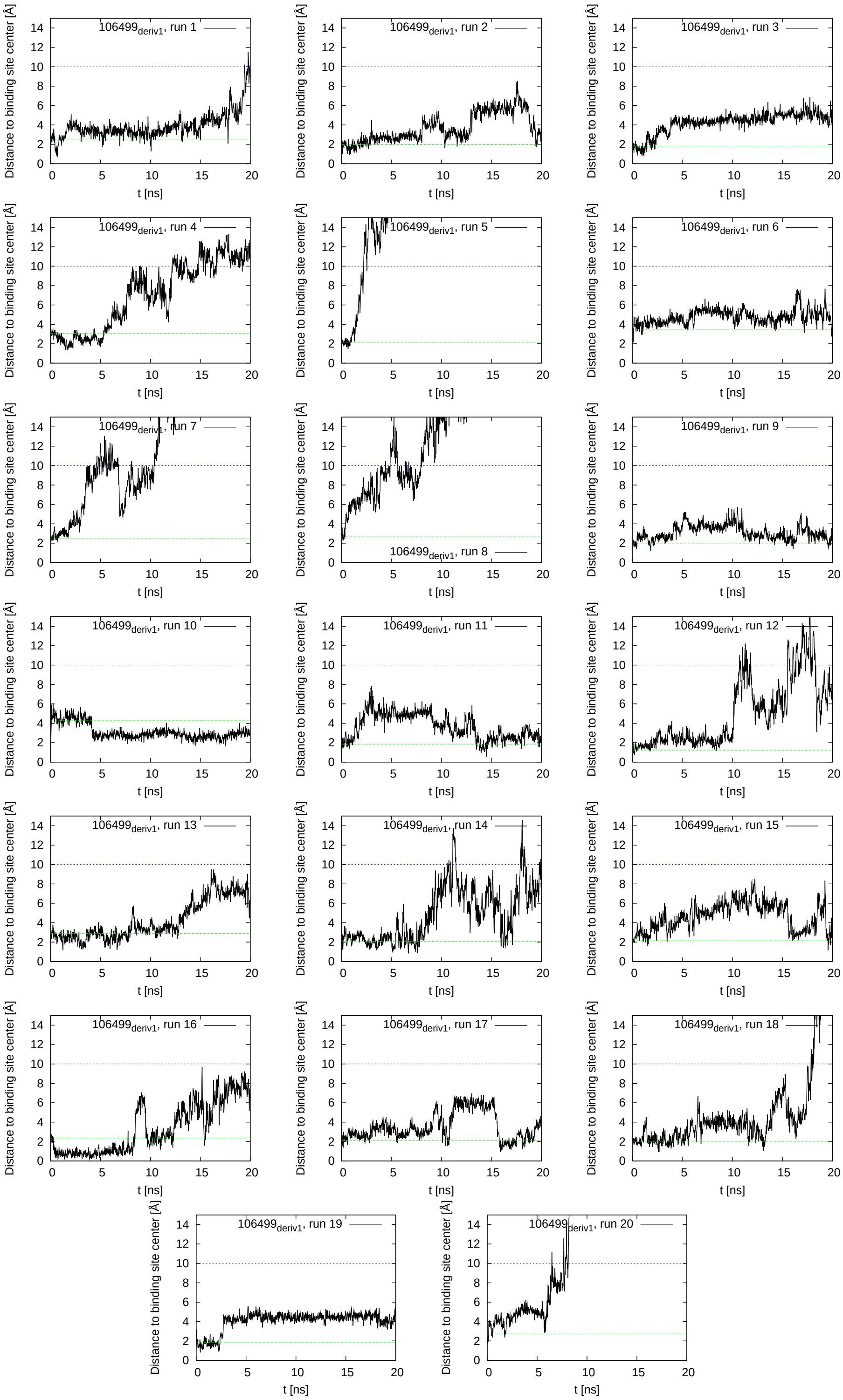


Figure S26: Time series of distance between centers of mass of compound 106499_{deriv1} and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

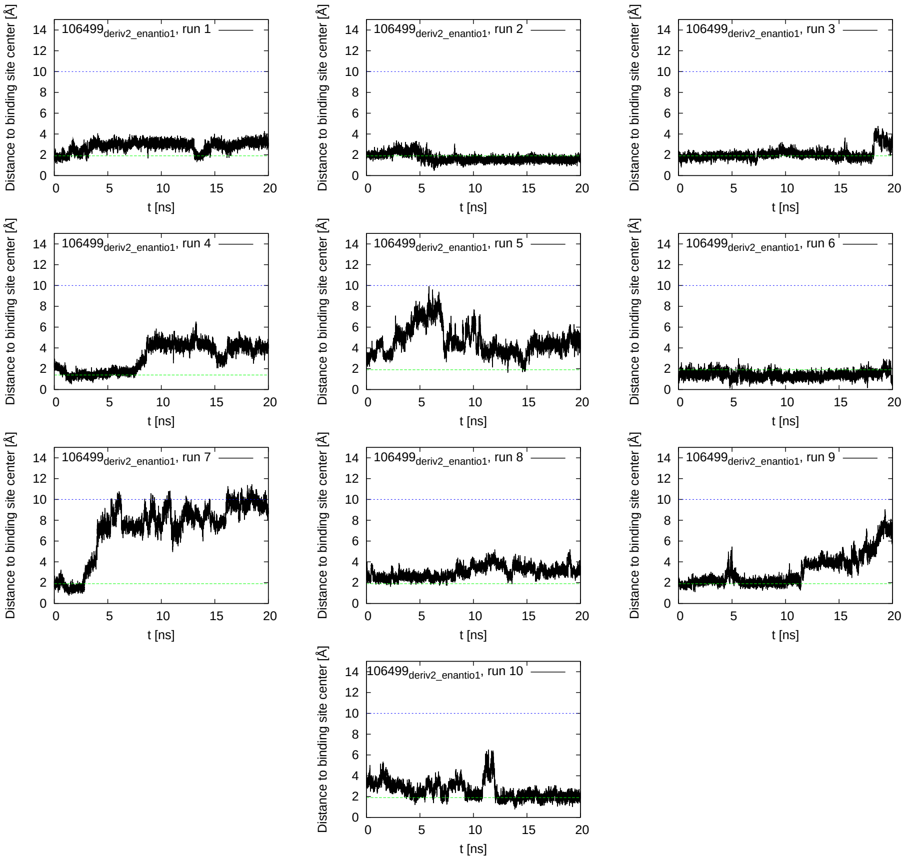


Figure S27: Time series of distance between centers of mass of compound $106499_{deriv2\_enantio1}$ and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help.

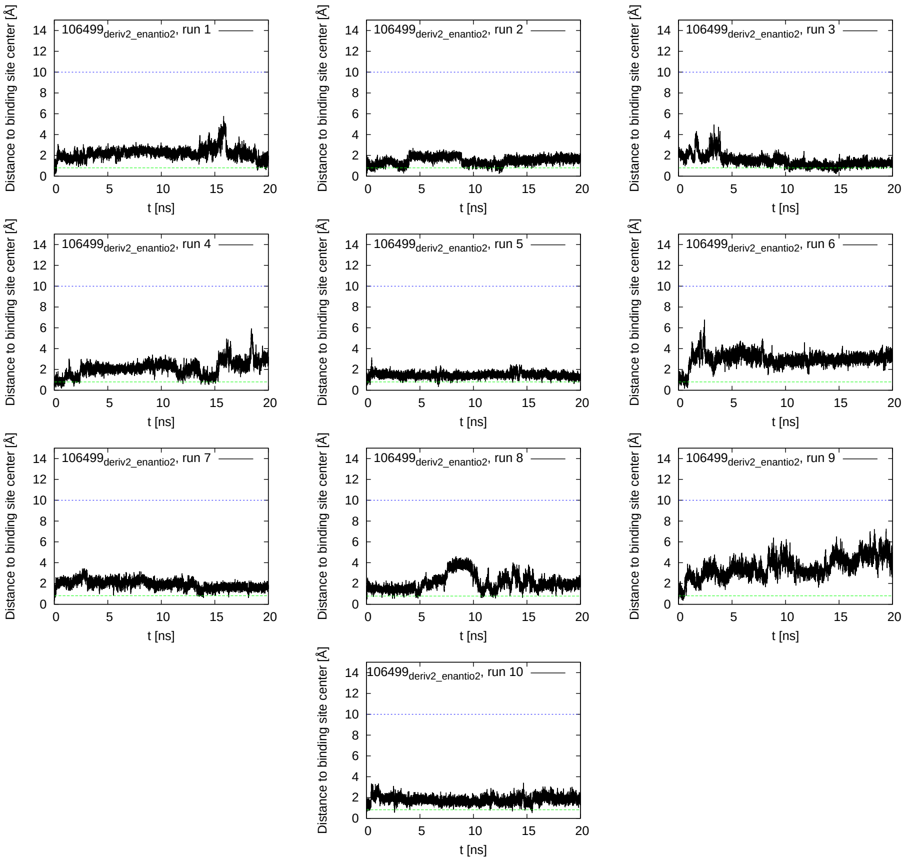


Figure S28: Time series of distance between centers of mass of compound $106499_{deriv2\_enantio2}$ and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help.

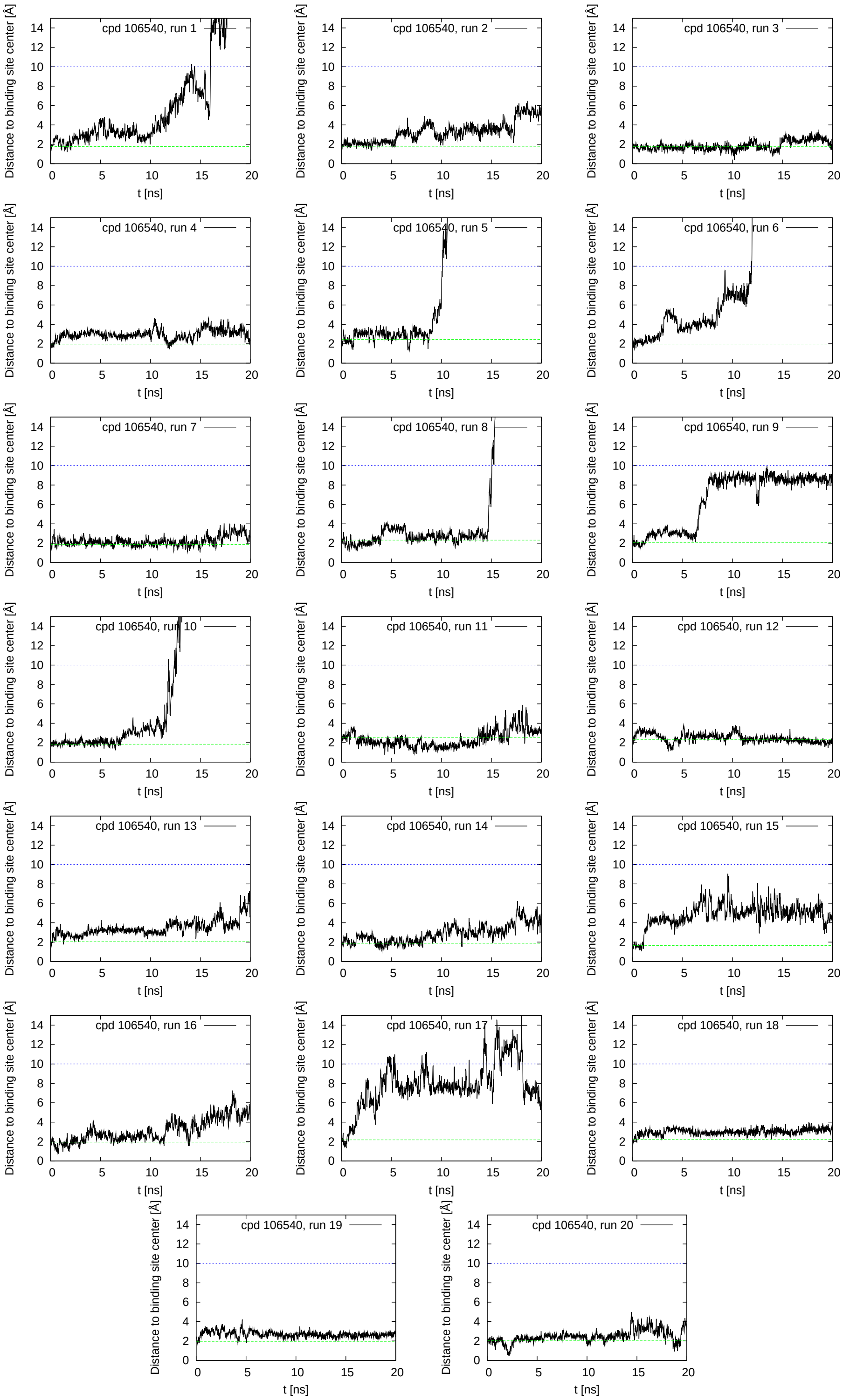


Figure S29: Time series of distance between centers of mass of compound 106540 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. "Compound" is abbreviated as "cpd" in the legends.

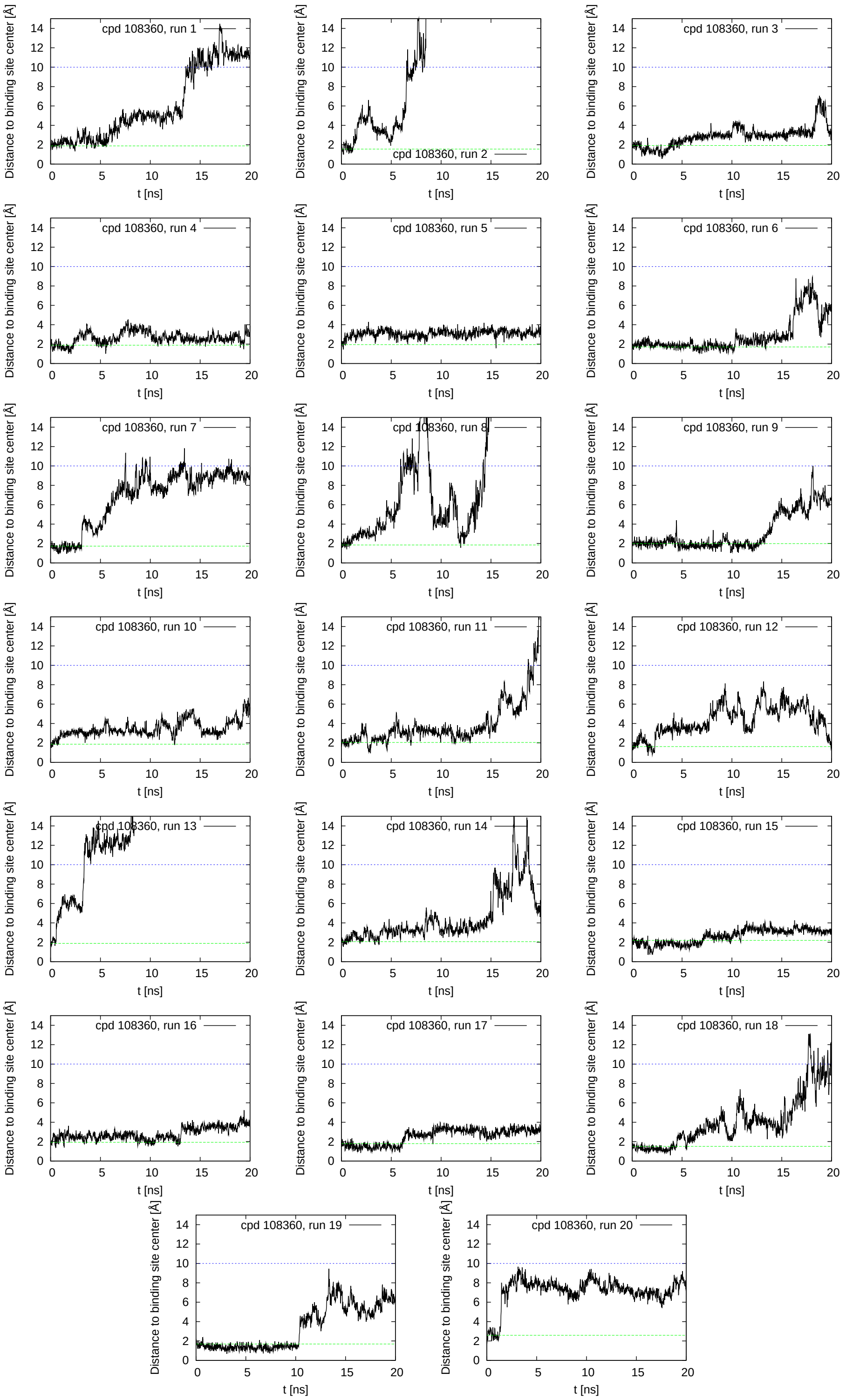


Figure S30: Time series of distance between centers of mass of compound 108360 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

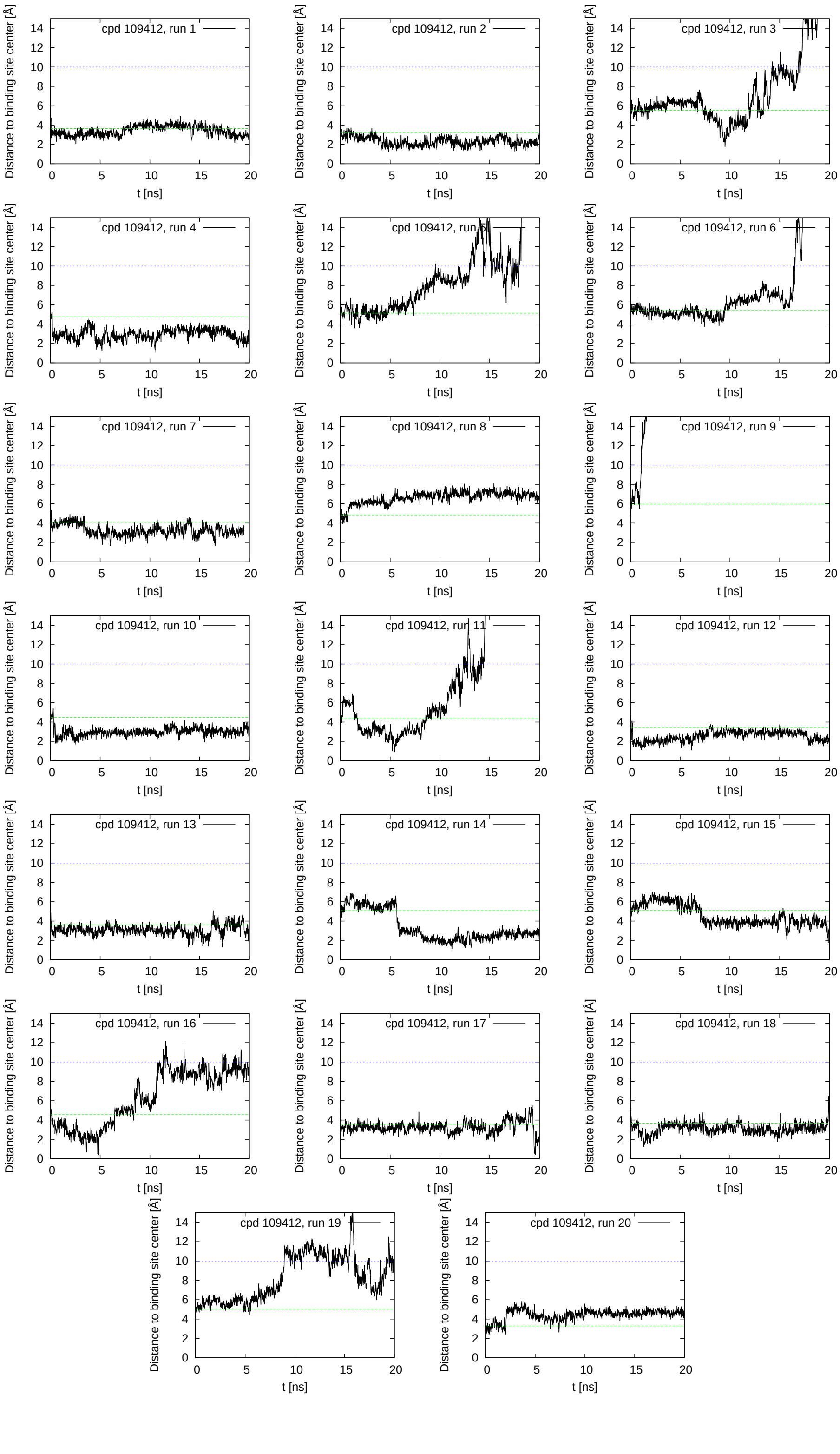


Figure S31: Time series of distance between centers of mass of compound 109412 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

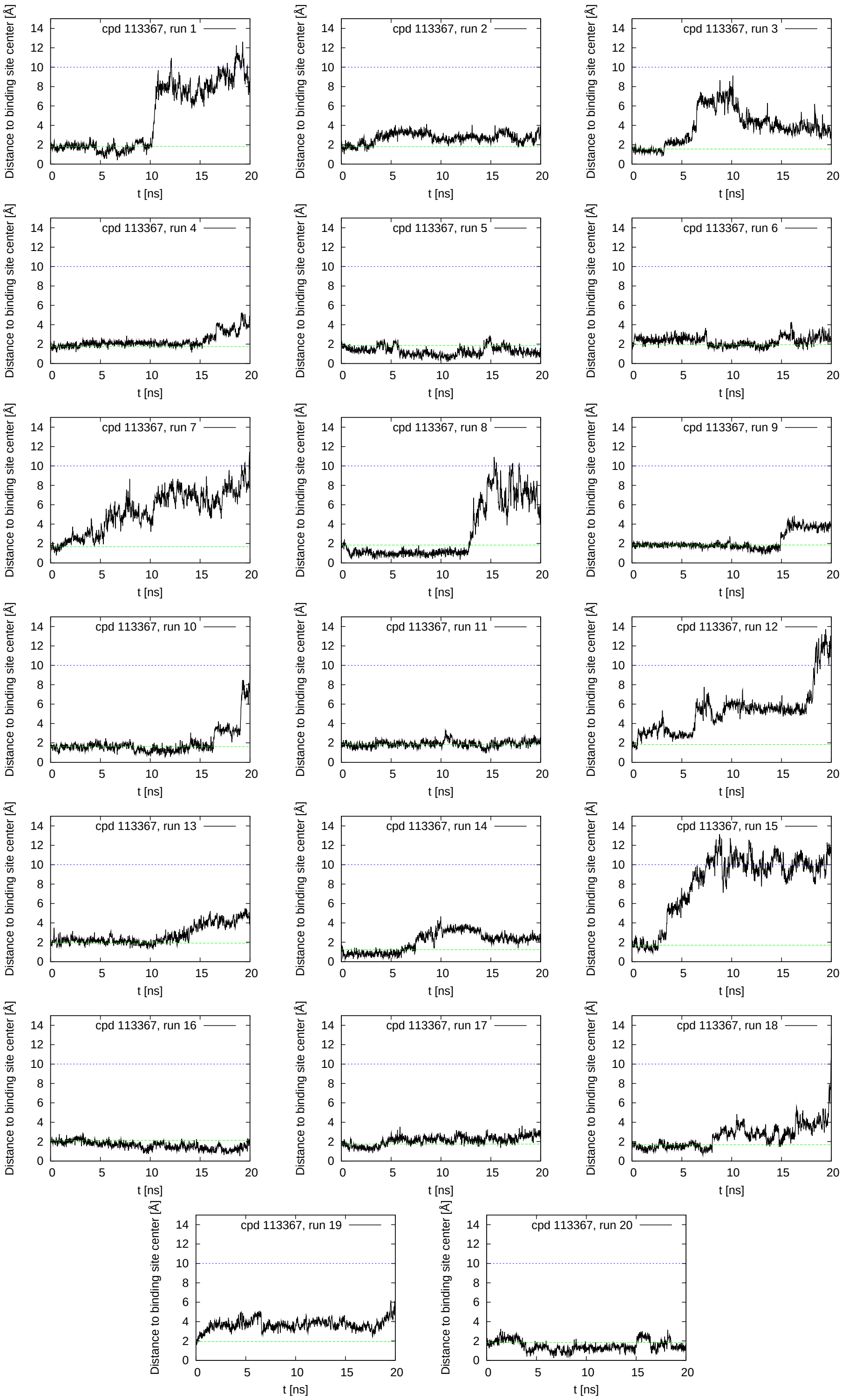


Figure S32: Time series of distance between centers of mass of compound 113367 and the actin binding site (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). The blue line illustrates the unbinding criterion of 10 Å, while the green line is the distance averaged over the first 10 snapshots and was drawn for visual help. “Compound” is abbreviated as “cpd” in the legends.

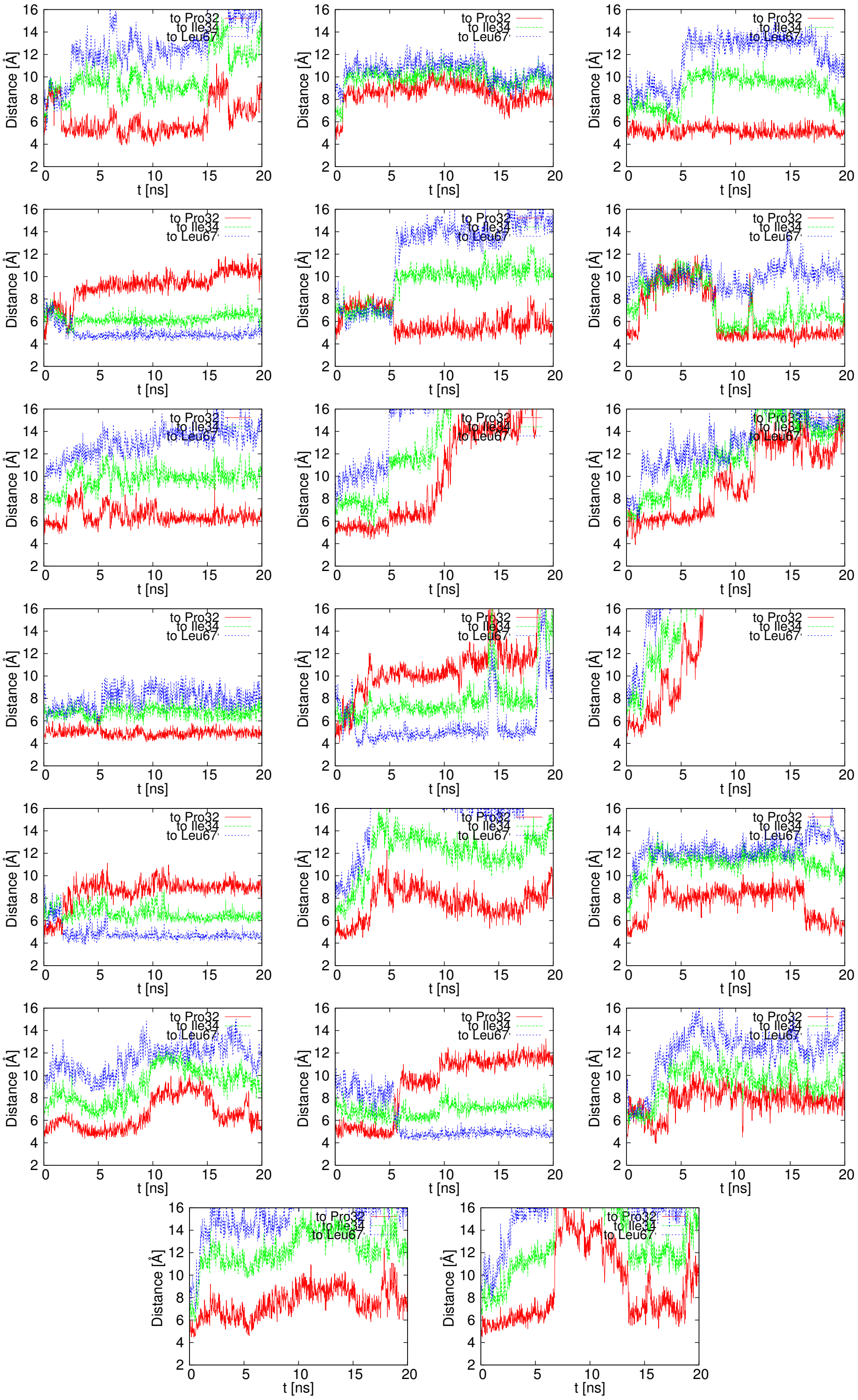


Figure S33: Time series of distance between isopropyl group of compound 42070 and hydrophobic residues (Pro32 (red), Ile34 (green), and Leu67 (blue)) in binding site

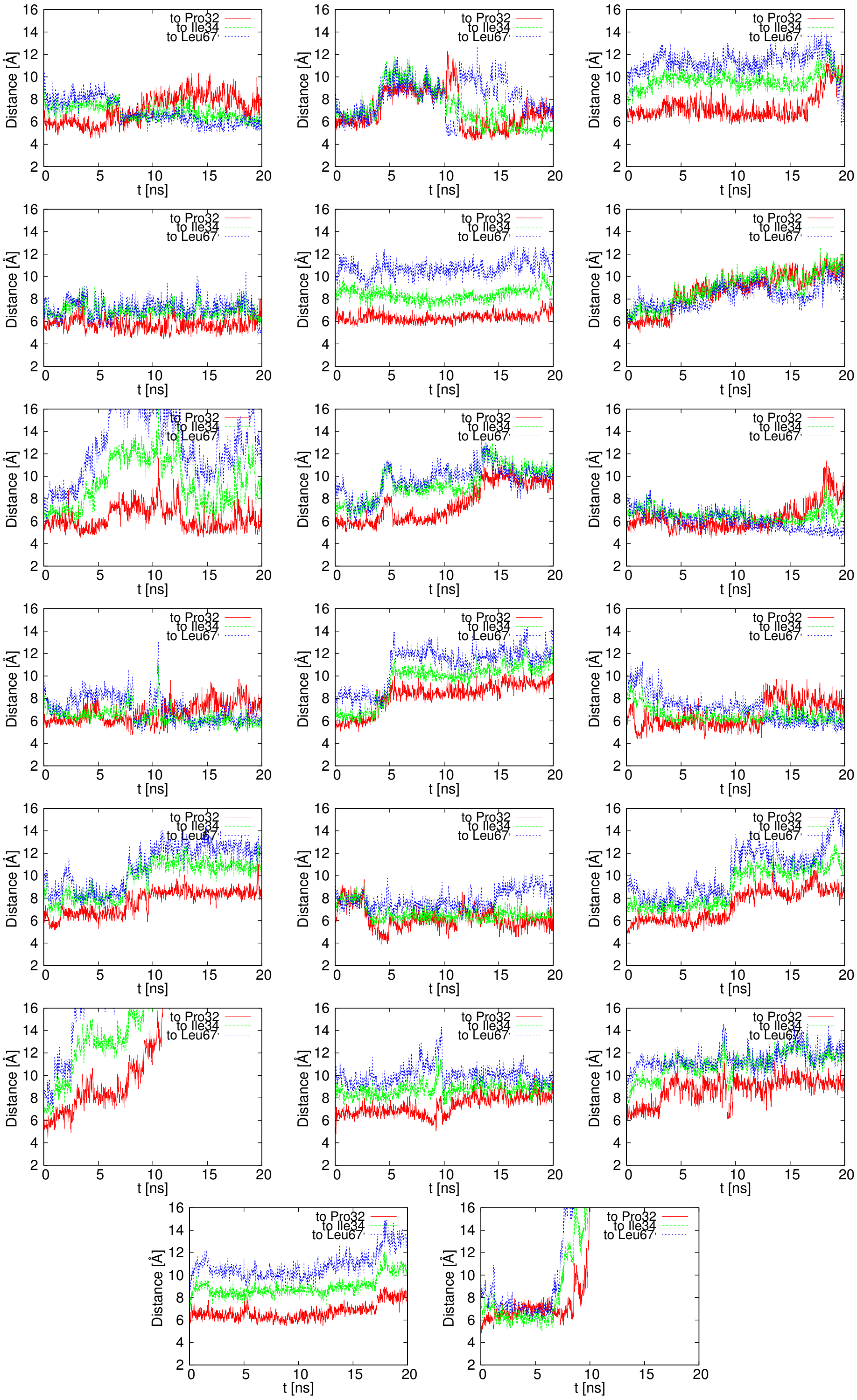


Figure S34: Time series of distance between isobutyl group of compound 42070_{opt1} and hydrophobic residues (Pro32 (red), Ile34 (green), and Leu67 (blue)) in binding site.

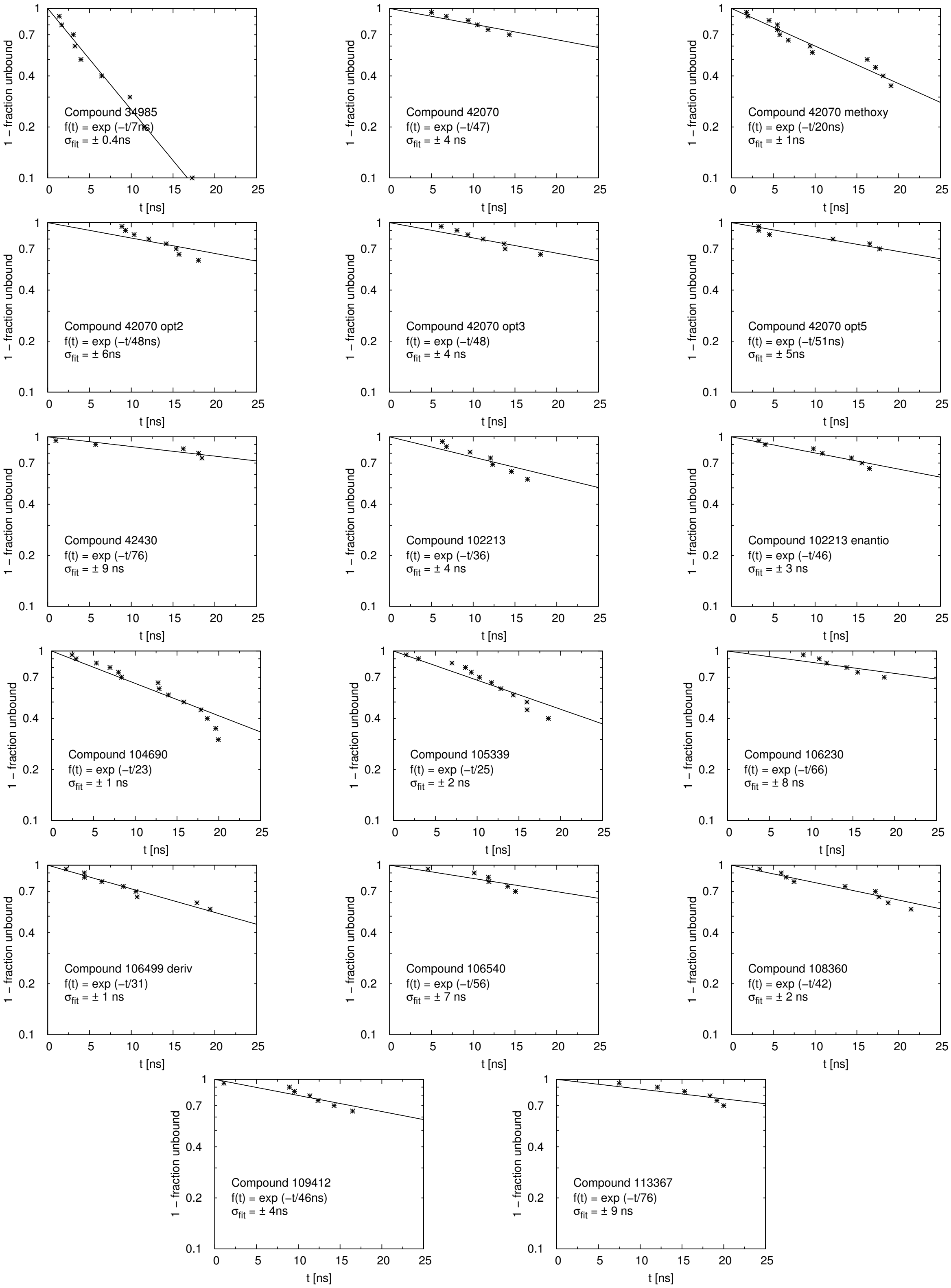


Figure S35: Single exponential kinetics of unbinding for ligands with at least five unbinding events. The plots show the cumulative distribution f(t) of the unbinding times observed in the 10-20 runs. The unbinding times and standard deviation of the fits (σ_{fit}) are given in the legends.

Chapter 3

De Novo design of G-actin inhibitors

Introduction

Computer-aided drug design is an established technology supporting experimental drug discovery techniques such as high-throughput screening in the search for novel pharmaceutically active agents. The most commonly used computational approach to drug discovery is virtual screening, where large libraries of existing compounds are docked into a protein target and their binding affinities estimated by a scoring function. An alternative approach to computer-aided drug discovery is *de novo* design, in which potential ligands are constructed "from scratch", allowing the design of truly novel chemical structures for drug development. [1, 2] Since its introduction about 20 years ago [1, 3], several programs for *de novo* design have been published such as GANDI [4], FlexNovo [5], MEGA [6], AutoGrow [7], and SQUIRREL$_{\text{novo}}$ [8]. Both receptor-based and ligand-based scoring functions are commonly employed, where the latter is based on a similarity measure to a compound known to bind to the target of interest. The simultaneous optimization of more than one scoring function is referred to as multiobjective optimization. In GANDI the individual scores of a force field energy term and the 2D or 3D similarity to a known inhibitor are combined either by a weighted-sum [9] or a Pareto-based [10] scoring approach.

The main advantage of *de novo* design over virtual screening is that a much larger part of chemical space can be sampled. However, due to the vastness of chemical space, which has been estimated to comprise appr. 10^{60}-10^{100} druglike molecules, it is not feasible to evaluate all theoretically possible compounds. [1] One of the main difficulties encountered in *de novo* design is the generation of synthetically accessible molecules. This issues has been addressed by using building blocks obtained by cleavage of existing compounds [11,12] and by applying connection rules derived from organic synthesis, such as the RECAP rules [13], facilitating a combination of building blocks by standard organic reactions.

Actin is a major constituent of the cytoskeleton and plays a crucial role in various cellular processes such as cell motility, adhesion, cytokinesis, and division. [14] Malignant modifications of the actin cytoskeleton are observed in tumor cells, making actin a potential drug target against cancer. [15–18] Additionally, actin-binding compounds are promising leads for the treatment of glaucoma by disrupting the actin cytoskeleton in the trabecular meshwork, a type of tissue in the eye, thereby increasing aqueous humour outflow and alleviating pressure on the optic nerve. [19] As of today the only known actin inhibitors binding to the so-called pointed end above the nucleotide binding site belong to the latrunculin family, a class of macrolide toxins produced by sponges in the Red Sea to fight predators. [20, 21] Due to the difficulties associated with the extraction from natural sources and the synthesis of latrunculins as well as other actin-binding natural products, there is a demand for novel chemical scaffolds for the development of actin-targeting anticancer and antiglaucoma drugs.

This study aims at identifying novel actin inhibitors targeting the same binding cleft as latrunculins by *de novo* design with GANDI. The crystal structure of the actin/latrunculin complex [21] shows the thiazolidinone ring of latrunculin buried in a subpocket deep in the binding site. As this unique moiety is shared by all active representatives of the latrunculin family, it is likely to be a key factor for a high binding affinity and selectivity. To identify potent actin inhibitors, the thiazolidinone ring is used as an anchor fragment for growing fragments obtained by decomposition and 2D structural clustering of a library of leadlike compounds. Different scoring schemes implemented in GANDI are evaluated to determine an optimal scoring method for the generation of novel actin-binding molecules. A number of interesting molecules presenting similar hydrogen bond patterns and more favorable van der Waals interaction efficiencies relative to latrunculin were obtained using GANDI, suggesting these molecules are promising candidates for chemical synthesis and measurement of *in vitro* binding affinity.

Materials and Methods

Protein preparation

The coordinates of actin in complex with latrunculin A [21] were downloaded from the PDB database (PDB ID: 1ESV). The inhibitor and all water molecules were removed. As residues 1-5 are missing in the X-ray structure, the -$COCH_3$ group was added to the N-terminal Thr6. The C-terminus was considered negatively charged. Hydrogens were added according to a pH of 7.4 using Babel [22], and CHARMm atom types [23] were assigned with Witnotp [24]. Hydrogen positions were subject to an energy minimization of 100 steps of the steepest descent (SD) and 1000 steps of the adopted basis Newton-Raphson algorithm using the program CHARMM [25, 26] and the CHARMm force field [23]. During the minimization a distance-dependent dielectric function of $\epsilon(r) = 4r$ was used, and all coordinates present in the X-ray structure were kept fixed.

Preparation of latrunculin as target for structural similarity

The coordinates of latrunculin were taken from the PDB structure 1ESV. Hydrogens were added according to a pH of 7.4 using Babel, and CHARMm atom types were assigned with Witnotp. Hydrogen positions were energy-minimized as described above for the protein.

Preparation of thiazolidinone ring as anchor fragment

The thiazolidinone ring was prepared by taking the coordinates from the crystallographic latrunculin in the PDB structure 1ESV. Subsequently, the structure was prepared as described before for latrunculin. To allow for a degree of conformational flexibility and increase the chances that GANDI finds solutions, the thiazolidinone ring was docked into the binding site using SEED [27, 27]. Poses docked into the same subpocket and with a similar orientation as the thiazolidinone ring in the crystal structure were selected as additional starting points for GANDI, yielding a total of six different poses.

Preparation and docking of fragment library

The leadlike subset (4.5 million compounds) of the April 2010 version of the ZINC database [28] was downloaded and decomposed into fragments by DAIM [29]. The fragments were clustered according to 2D structural similarity using DAIM fingerprints and a Tanimoto threshold of 0.98. A small diversity set of building blocks is preferable to a large library due to a strong increase in computation time by combinatorial explosion. The 16500 cluster representatives obtained by DAIM were subject to a CHARMM minimization of 100 steps of the steepest descent (SD) and 5000 steps of the adopted basis Newton-Raphson algorithm, using a distance-dependent dielectric function of $\epsilon(r) = 4r$ and the CHARMm force field. Subsequently, the fragments were docked into the binding site using SEED, keeping only poses with a SEED energy more favorable than 0 kcal/mol.

De novo design with GANDI

Version 2.0 of GANDI [4] (Genetic Algorithm-Based de Novo Design of Inhibitors) was used to join the fragments previously docked by SEED. GANDI employs a parallel genetic algorithm in which a gene value corresponds to a single docked fragment position read in. Individuals of both the parent and the children population compete for survival. In this study, the fragments were successively grown to the thiazolidinone anchor fragment. Four islands, 100 individuals per island, and 1000 iteration steps were used, and each run was repeated three times with different random seeds. Between 100-300 molecules were generated in a single run.

The total score is minimized until the maximum number of iterations is reached. Two different scoring schemes are available. The weighted-sum approach employs a linear combination of scoring function terms:

$$S_{\mathrm{total}} = w_{\mathrm{ff}} E_{\mathrm{ff}} - w_{\mathrm{3D}} \mathrm{Sim}_{\mathrm{3D}} - w_{\mathrm{PH4}} \mathrm{Sim}_{\mathrm{PH4}} \tag{3.1}$$

where S_{total} is the total score, E_{ff} the force field energy, $\mathrm{Sim}_{\mathrm{3D}}$ the 3D similarity to a known ligand, and $\mathrm{Sim}_{\mathrm{PH4}}$ the pharmacophore score. The weights for the scoring function terms are w_{ff}, w_{3D}, and w_{PH4}, respectively. The force field energy consists of inter- and intramolecular van der Waals and electrostatic terms. [4]

The 2D similarity term implemented in GANDI was not considered because using the knowledge of the binding mode of latrunculin was preferred to a ligand-based design. For the pharmacophore scoring term, the presence of a hydrogen bond to residues Tyr69 and Arg210 was evaluated (see Results section), using a distance criterion of 2.8 Å between hydrogen and acceptor atoms. As a basis of comparison, the longest hydrogen bond distance in latrunculin is 2.6 Å between the oxygen atom of latrunculin's hydroxyl group and the closest hydrogen of the guanidinium group of Arg210. A slightly higher distance criterion compared to the longest bond in the actin/latrunculin complex was chosen to allow for some side chain flexibility, considering the usage of a rigid protein in GANDI.

In the Pareto-based scoring an individual A dominates over another individual B if the scores of all scoring functions of A are equal or lower (at least one) than the corresponding scores of B. A Pareto-based scoring does not require the tuning of any scoring function weights.

Table 3.1 gives an overview of the different setups used in this study.

setup	scoring approach[a]	PH4 as scoring term or filter[b]	w_{ff}	w_{3D}	w_{PH4}	no. of fragments to link[c]	linker set[d]	Presence of OH group at anchor[e]
1	ws	sc	0.02	1	1	2	small	No
2	ws	f	0.02	1	-	2	small	No
3	p	sc	-	-	-	2	small	No
4	p	f	-	-	-	2	small	No
5	ws	f	1	1	-	2	small	No
6	ws	f	0.01	1	-	2	small	No
7	ws	f	1	1	-	2	large	No
8	ws	f	1	1	-	3	large	No
9	ws	f	1	1	-	2	large	Yes

Table 3.1: Overview of used GANDI setups.
[a] The weighted-sum approach is labeled "ws", Pareto-based scoring "p".
[b] The pharmacophore term (abbreviated as "PH4") is either used in the scoring function ("sc") or as a filter criterion ("f") after scoring.
[c] In all runs, one fragment is the thiazolidinone anchor fragment to which either two fragments (in setup 8) or one fragment (in the remaining setups) are grown.
[d] The two linker sets are explained in the text.
[e] In setup 9, the coordinates of the hydroxyl group of latrunculin are included in the anchor fragment.

Two different sets of linkers were used. The first set is the linker library from the GANDI testcase, which consists of small molecules like ethanol, acetamide, acetate etc. This linker set is used to produce compounds with a low molecular weight, and will be referred to as the "small" linker library. As a second set of linkers the fragment library was filtered for fragments with a molecular weight below 150 g/mol, yielding about 2100 molecules. This linker set will be referred to as the "large" linker library to distinguish it from the small linker set. All heavy atom - hydrogen atom vectors in fragments and linkers were used as connection vectors. In setups where the pharmacophore term is not used in the scoring function, but as a filter criterion after scoring (setups 2 and 4-9), the presence of a hydrogen bond to either Tyr69 or Arg210 was set as mandatory.

All fragments of the docked fragment library were used in the different setups, with the exception of setup 7 (only fragments with a molecular weight equal to or above 200 g/mol) and 8 (only fragments with a molecular weight below 200 g/mol) to obtain more reasonably sized compounds. The volume explored by GANDI was restricted to that within 2 Å of latrunculin in its bound conformation.

The number of GANDI-generated molecules forming hydrogen bonds to the receptor was measured with CHARMM using a distance criterion of 2.8 Å between hydrogen and acceptor atoms. The CHARMm protein-compound total interaction energy and protein-compound van der Waals interaction efficiency, i.e., the energy divided by the number of heavy atoms, were calculated after minimizing the compounds inside the rigid protein using 100 steps of the steepest descent (SD) and 5000 steps of the adopted basis Newton-Raphson algorithm in CHARMM. A distance-dependent dielectric function of $\epsilon(r)=4r$ and default nonbonding cutoffs were used in the minimization.

Results

In the actin/latrunculin complex, latrunculin is involved in hydrogen bonds to Tyr69, Arg210, Asp157, and Thr186, where the latter two residues interact with the thiazolidinone moiety used as anchor fragment in GANDI (Fig. 3.1). The two residues Tyr69 and Arg210, which do not interact with the anchor fragment, will be referred to as "key" residues in the following sections. Their participation in strong hydrogen bonds is likely to be crucial for a high binding affinity. The guanidinium group of Arg210 can function as a strong donor, and Tyr69 is placed relatively deep in the binding site, allowing it to participate in strong hydrogen bonds due to lower dielectric shielding. A desolvation of Tyr69 upon ligand binding without forming a new hydrogen bond is likely to be energetically unfavorable.

To illustrate the minimization of the total score in a GANDI run, the evolution of individual scoring function terms and the total score is shown for two individuals of setup 2 as an example (Fig. 3.2). The total score (Fig. 3.2, top left) decreases monotonically over the 1000 performed iterations. For iteration steps where the total score remains constant relative to the previous step, no better solution was found by the program and the currently selected linkers and docked fragments remain the same. No new molecules are introduced into the population until individuals with a lower total score surviving the selection are found. Both the force field energy and the 3D similarity (Fig. 3.2, top right and bottom) can either decrease or increase because it is the weighted-sum of these terms which is optimized in the GANDI run. An increase in one of the two terms is always accompanied by a decrease in the other, resulting in an overall reduction of the total score. The evolution of the force field energy and the 3D similarity is substantially different for the two individuals. Since a different set of docked poses was assigned to the individuals at the start of the run, the optimization proceeds differently, i.e., different linkers and poses survive the selection after each iteration step, resulting in the generation of two different molecules.

Selection of optimal scoring method

Different scoring schemes implemented in GANDI were tested to determine an optimal scoring method for the generation of candidate ligands targeting actin. The quality of the binding modes generated by setups 1-4 (see Materials and Methods section) was evaluated according to the presence of hydrogen bonds to the key residues in the actin binding site and by visual examination. Both the weighted-sum and Pareto-based approach including a pharmacophore term in the scoring (setups 1 and 3) produce only around 20-25% of molecules with hydrogen bonds to Tyr69 or Arg210 (Fig. 3.3). The percentage of compounds with hydrogen bond interactions to Arg210 and especially to Tyr69 is significantly increased by using the pharmacophore function not in the scoring, but as a filter criterion (setups 2 and 4). This observation is plausible: When a pharmacophore term is incorporated in the scoring, its influence on the total score is lower than when it is used as an independent filter term alter scoring, due to the presence of the force field and 3D similarity terms. The percentage of compounds involved in hydrogen bonds to both Tyr69 and Arg210 is 4% for setup 1, 14% for setup 2, 5% for setup 3, and 18% for setup 4. Using the presence of both hydrogen bonds as a scoring function term or as a filter is a too strict criterion and leads to GANDI not finding many solutions. Since visual examination revealed that setup 2 results in the highest number of compounds with a reasonable filling of the binding site (e.g. by forming hydrophobic contacts to apolar side chains of Pro32 and Ile34), setup 2, i.e., a weighted-sum scoring with pharmacophore constraints as filter criterion, was selected as scoring method. Employing the chosen scoring scheme, further GANDI runs were performed with two different linker libraries and different weights for the scoring function terms to obtain a diverse set of compounds (setups 2 and 5-9).

Physicochemical properties of generated molecules

Fig. 3.4 shows the physicochemical properties of the compounds obtained by setups 2 and 5-9. Most compounds generated with a small linker library (Fig. 3.4, left panels) as well as those from setups 7 and 9 (Fig. 3.4, magenta and black lines in right panels) fulfill Lipinski's rule of five. [30] Growing two fragments with large linkers to the thiazolidinone moiety without the hydroxyl group (setup 7) leads to compounds which are on average 50 g/mol heavier than those obtained from setup 9. The hydroxyl group fills up space in the binding pocket, reducing the possibilities of placing big fragments. Only the combination of three fragments results in a high number of molecules with a too high molecular weight according to Lipinski's rules of five (setup 8, cyan line in Fig. 3.4). Appr. one third of the generated compounds have too many rotatable bonds as well (Fig. 3.4, (b)), which is not observed for the other GANDI setups. Due to their large size and high number of rotatable bonds and acceptors, some molecules obtained from setup 8 might not be appropriate for lead optimization into an orally bioavailable drug. More importantly, visual examination revealed that some of these molecules do not fill the space of the binding site as perfectly as the smaller molecules generated by combining two fragments. Since GANDI is not capable of forming ring systems, the generated molecules are linear. When connecting three fragments, the third fragment is placed at the "edge" of the binding site, preventing the formation of interactions.

Most compounds created by all GANDI setups conform to the Lipinski rules regarding clogP and the number of hydrogen bond donors and acceptors, with the exception of setup 8 with a significant amount of molecules with too many acceptors.

Even though it is informative to calculate the physicochemical properties of the created molecules, Lipinski's rule of five should not be used as a strict cutoff for the choice of candidate ligands. There are a number of drugs on the market which do not comply with the Lipinski rules, such as the immunosuppressant cyclosporine with a molecular weight of 1203 g/mol or the tyrosine kinase inhibitor nilotinib with a molecular weight of 530 g/mol for the treatment of chronic myelogenous leukemia, which are still commonly used therapeutic agents.

Evaluation of binding poses

To assess the quality of the binding poses generated by GANDI, the poses were evaluated according to the presence of hydrogen bond interactions to key residues. Additionally, the CHARMm total interaction energy and the CHARMm van der Waals interaction efficiency were computed after an energy minimization of the compounds within the protein to obtain a more reliable estimation of interaction energies. In contrast, the GANDI force field energy is calculated without a prior minimization. As shown by the set of compounds generated using the small linker library, different weights for the force field and 3D similarity term in the scoring function do not lead to significant differences in the number of compounds forming hydrogen bonds to key residues (Fig. 3.5, left panel). The small increase in number of compounds with interactions to Tyr69 using a low weight for the force field term (setup 6) is compensated by a decrease of interactions to Arg210. The number of molecules with hydrogen bonds to both key residues is around 10% for the three setups. Using the large linker library produces appr. 10% less molecules with interactions to Tyr69 (Fig. 3.5, right panel). Smaller linkers facilitate the design of compounds with a better fit within the binding site capable of interacting with the relatively buried Tyr69. In comparison to setup 7, the combination of three fragments in setup 8 produces a higher number of molecules interacting with Arg210 on account of the higher number of hydrogen bond acceptors in the created compounds (see previous section). Setup 9 results in the by far highest number of molecules with hydrogen bonds to both residues. The percentage of compounds bound to Arg210 is 100% for this setup due to the presence of latrunculin's hydroxyl group in the anchor fragment. The choice of linker library, number of fragments to connect, and chemical structure of the fragment used as anchor has a greater influence on the hydrogen bond forming properties of the resulting compounds than the weights of the scoring function terms.

Almost all generated compounds have a CHARMm total interaction energy significantly less favorable than latrunculin (Fig 3.6, (a)), which is not surprising given the low micromolar binding affinity of this natural product. A relatively high number of molecules with a lower interaction energy than latrunculin is only found for setup 8. This result is to be expected as the CHARMm total interaction energy is highly size-dependent, mostly due to the van der Waals contribution, and most molecules produced by setup 8 are heavier than latrunculin which has a molecular weight of 422 g/mol. To address the issue of size-dependency, the CHARMm van der Waals interaction efficiency was computed to evaluate the strength of van der Waals interactions. Each of the run setups generated a small number of compounds with more favorable van der Waals interaction efficiencies than latrunculin (Fig. 3.6, (b)). For the large linker library, setup 9 achieves the most favorable and setup 8 the most unfavorable CHARMm van der Waals efficiencies, which is closely related to the different distributions of molecular weights.

The most promising compounds satisfy both criteria described above, i.e., hydrogen bonds to key residues and a favorable CHARMm van der Waals interaction efficiency. The number of compounds with interactions to both Tyr69 and Arg210 and a CHARMm van der Waals efficiency more favorable than -1.70 kcal/mol (latrunculin: -1.75 kcal/mol) are 3.0% (setup 2), 2.8% (setup 5), 3.0% (setup 6), 0.3% (setup 7), 0.3% (setup 8), and 12.1% (setup 9). The chemical structures of two molecules of setup 2 and one molecule of setup 9 are shown in Fig. 3.7 as an example. Setup 9 results in the highest number of compounds fulfilling both criteria, caused mainly by the hydroxyl group at thiazolidinone ring which substantially increases the number of hydrogen bonds to both key residues. The higher number of compounds satisfying both criteria for GANDI setups 2, 5, and 6 compared to 7 and 8 results mainly from more favorable CHARMm van der Waals efficiencies caused by lower molecular weights of the created compounds.

Conclusion

In this study the *de novo* design software GANDI was used to design novel actin inhibitors. Based on both the ability to form specific hydrogen bonds and a reasonable fit within the binding site, the weighted-sum approach was found to produce the highest quality of binding poses, and was therefore chosen as scoring scheme. No significant differences in the quality of binding modes were observed by adjusting the weights of the force field and 3D similarity term in the scoring. Even though notable differences in both hydrogen bond forming properties and CHARMm interaction energies were obtained for compounds generated using different scoring function terms and linker sets, a close visual examination revealed reasonable binding poses for all six run setups. While setups 2, 5, 6, and 7 result in molecules which fit well into the binding site, the number of compounds involved in both key hydrogen bonds is highest for setup 9 due to the presence of latrunculin's OH group. Visual examination revealed a number of interesting structures for setup 8 as well which combines three instead of two fragments, though several compounds do not fit perfectly into the binding site. The most promising candidate ligands were obtained by selecting molecules with a favorable CHARMm van der Waals efficiency interacting with both key residues, combined with a visual examination of the binding poses. Applying this procedure, interesting novel structures from six GANDI setups are identified for chemical synthesis and measurement of *in vitro* binding affinity to actin. In addition to the fragments obtained from the ZINC library of commercially available compounds, we are currently docking a library of natural product-derived fragments for use in GANDI. A natural product-inspired fragment library is expected to be enriched in structural motifs favorable for binding to proteins, and molecules constructed from these fragments are expected to be interesting candidate ligands for the design of actin inhibitors.

Bibliography

[1] Schneider G, Fechner U. Computer-based *de novo* design of drug-like molecules. Nature Rev Drug Discov 2005;4:649–663.

[2] Mauser H, Guba W. Recent developments in *de novo* design and scaffold hopping. Curr Opin Drug Discov Devel 2008;11:365–374.

[3] Danziger DJ, Dean PM. Automated site-directed drug design: a general algorithm for knowledge acquisition about hydrogen-bonding regions at protein surfaces. Proc R Soc Lond B Biol Sci 1989;236:101–113.

[4] Dey F, Caflisch A. Fragment-based *de Novo* ligand design by multiobjective evolutionary optimization. J Chem Inf Model 2008;48:679–690.

[5] Degen J, Rarey M. FlexNovo: structure-based searching in large fragment spaces. ChemMedChem 2006;1:854–868.

[6] Nicolaou CA, Apostolakis J, Pattichis CS. *De novo* drug design using multiobjective evolutionary graphs. J Chem Inf Model 2009;49:295–307.

[7] Durrant JD, Amaro RE, McCammon JA. Autogrow: a novel algorithm for protein inhibitor design. Chem Biol Drug Des 2009;73:168–178.

[8] Proschak E, Zettl H, Tanrikulu Y, Weisel M, Kriegl JM, Rau O, Schubert-Zsilavecz M, Schneider G. From molecular shape to potent bioactive agents I: bioisosteric replacement of molecular fragments. ChemMedChem 2009;4:41–44.

[9] Pegg SC, Haresco JJ, Kuntz ID. A genetic algorithm for structure-based *de novo* design. J Comput-Aided Mol Des 2001;15:911–933.

[10] Nicolaou CA, Apostolakis J, Pattichis CS. *De Novo* drug design using multiobjective evolutionary graphs. J Chem Inf Model 2009;49:295–307.

[11] Gillet VJ, Myatt G, Zsoldos Z, Johnson AP. SPROUT, HIPPO and CAESA: Tools for *de novo* structure generation and estimation of synthetic accessibility. Perspect Drug Discov Des 1995;3:34–50.

[12] Douguet D, Munier-Lehmann H, Labesse G, Pochet S. LEA3D: A computer-aided ligand design for structure-based drug design. J Med Chem 2005;48:2457–2468.

[13] Lewell X, Judd D, Watson S, Hann M. RECAP-Retrosynthetic Combinatorial Analysis Procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. J Chem Inf Model 1998;38:511–522.

[14] Schmidt A, Hall MN. Signaling to the actin cytoskeleton. Annu Rev Cell Dev Biol 1998; 14:305–338.

[15] Jordan MA, Wilson L. Microtubules and actin filaments: Dynamic targets for cancer chemotherapy. Curr Opin Cell Biol 1998;10:123–130.

[16] Janmey PA, Chaponnier C. Medical aspects of the actin cytoskeleton. Curr Opin Cell Biol 1995;7:111–117.

[17] Giganti A, Friederich E. The actin cytoskeleton as a therapeutic target: state of the art and future directions. Prog Cell Cycle Res 2003;5:511–525.

[18] Lambrechts A, Van Troys M, Ampe C. The actin cytoskeleton in normal and pathological cell motility. Int J Biochem Cell Biol 2004;36:1890–1909.

[19] Peterson JA, Tian B, Geiger B, Kaufman PL. Effect of latrunculin B on outflow facility in monkeys. Exp Eye Res 2000;70:307–313.

[20] Nèeman I, Fishelson L, Kashman Y. Isolation of a new toxin from the sponge *Latrunculia magnifica* in the Gulf of Aquaba (Red Sea). Marine Biol 1975;30:293–296.

[21] Morton WM, Ayscough KR, McLaughlin PJ. Latrunculin alters the actin-monomer subunit interface to prevent polymerization. Nat Cell Biol 2000;2:376–378.

[22] O'Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: An open chemical toolbox. J Cheminf 2011;3:33–47.

[23] Momany FA, Rone R. Validation of the general purpose QUANTA3.2/CHARMm force field. J Comput Chem 1992;13:888–900.

[24] Widmer A. Witnotp: A computer program for molecular modeling. 1997; Novartis: Basel.

[25] Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.

[26] Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao

J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: The biomolecular simulation program. J Comput Chem 2009; 30:1545–1614.

[27] Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A. Exhaustive docking of molecular fragments on protein binding sites with electrostatic solvation. Proteins: Struct Func Bioinf 1999;37:88–105.

[28] Irwin JJ, Shoichet BK. ZINC - a free database of commercially available compounds for virtual screening. J Chem Inf Model 2005;45:177–182.

[29] Kolb P, Caflisch A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment- based high-throughput docking. J Med Chem 2006;49:7384–7392.

[30] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev 2001;46:3–26.

Tables and Figures

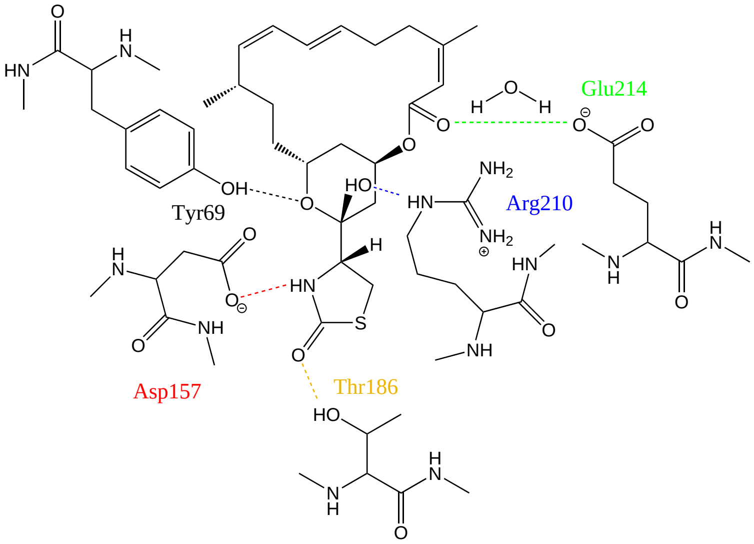


Figure 3.1: Two-dimensional scheme of latrunculin and G-actin residues involved in hydrogen bonds according to the 1ESV crystal structure. The thiazolidinone moiety used as anchor fragment in GANDI forms hydrogen bonds to Asp157 and Thr186. The hydrogen bond to Glu214 is water-bridged and was therefore not considered in this study.

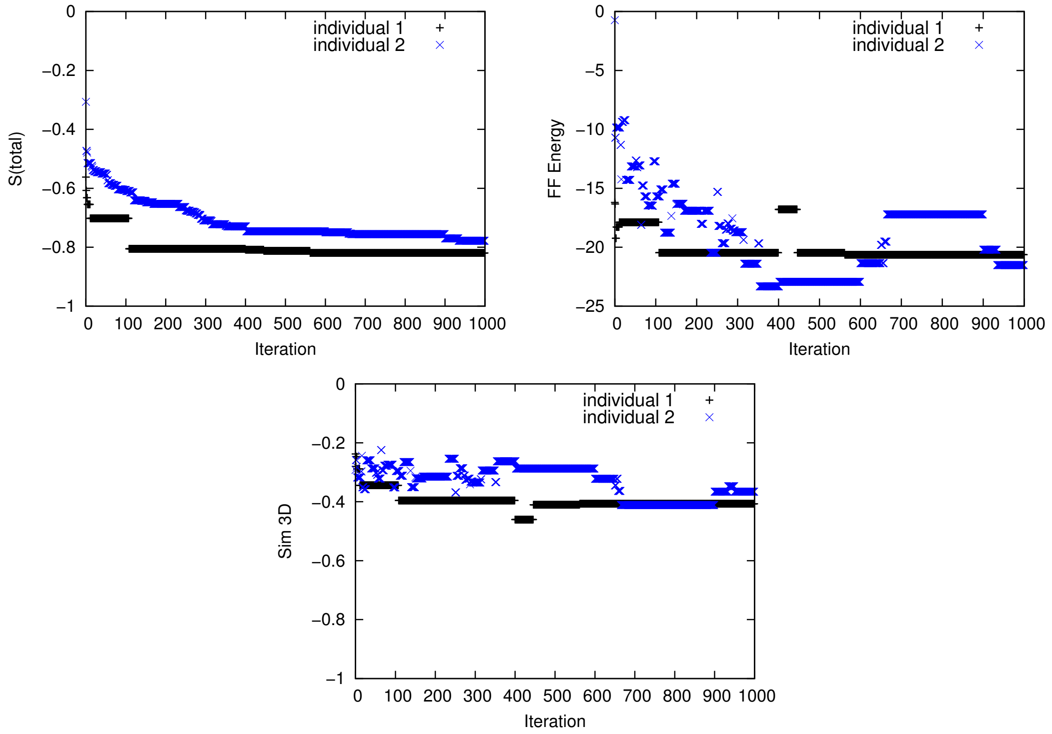


Figure 3.2: Evolution of scoring function terms for two individuals generated by GANDI setup 2 (black: first individual in island 1 of run 1, blue: first individual in island 4 of run 1). (Top left) Evolution of total score. (Top right) Evolution of force field energy. (Bottom) Evolution of 3D similarity to latrunculin.

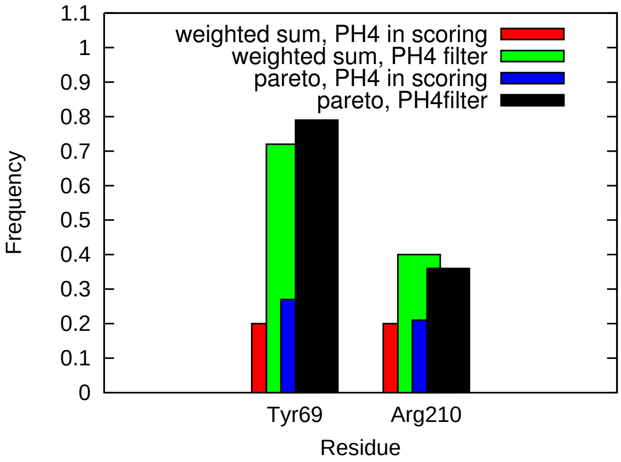


Figure 3.3: Percentage of generated molecules forming hydrogen bonds to key residues Tyr69 and Arg210 for GANDI setups 1 (red), 2 (green), 3 (blue), and 4 (black). The scoring schemes are indicated in the legend. The term “pharmacophore” is abbreviated as “PH4”.

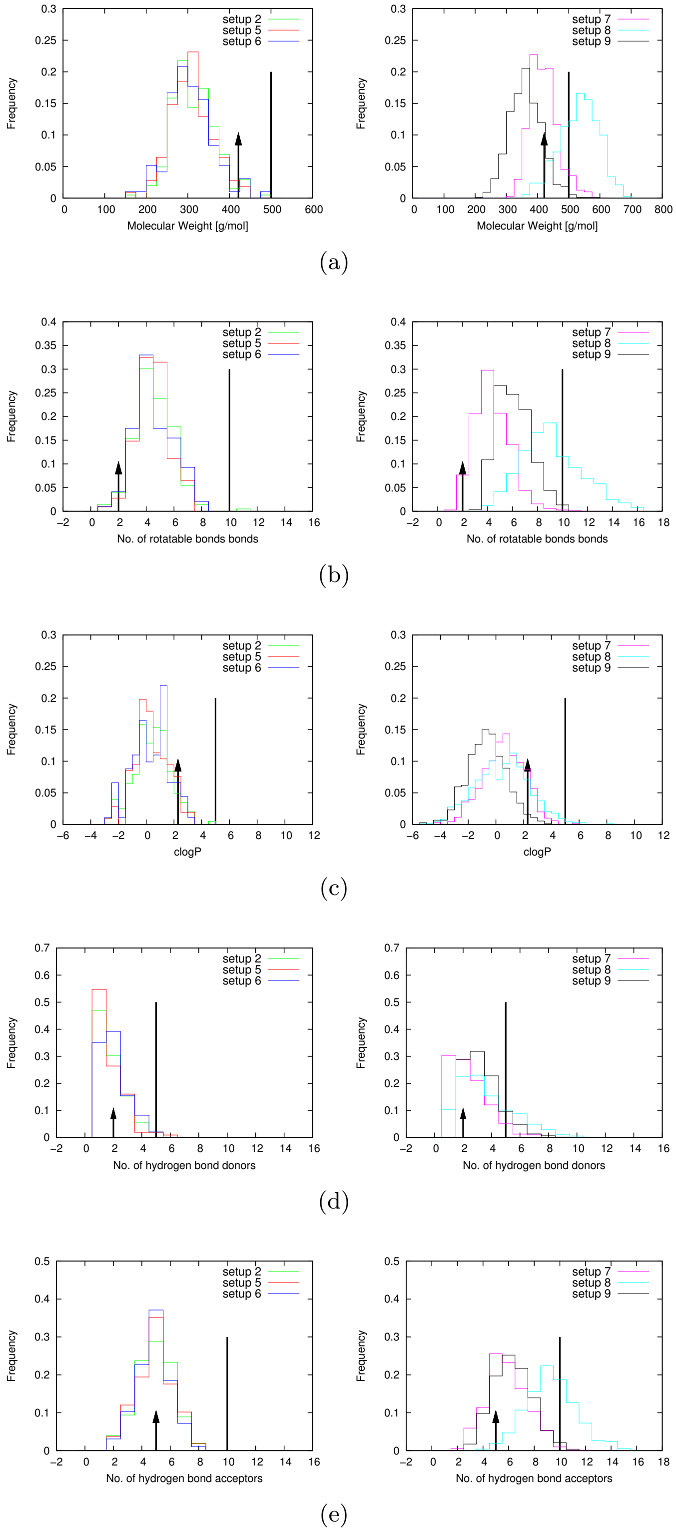


Figure 3.4: Physicochemical properties of compounds generated using GANDI setups 2, 5, and 6 (left) and 7-9 (right). (a) Distribution of molecular weights. (b) Distribution of number of rotatable bonds. (c) Distribution of clogP. (d) Distribution of number of hydrogen bond donors. (e) Distribution of number of hydrogen acceptors. The vertical black lines represent the thresholds defined by Lipinski's rule of five. [30] For comparison, the values for latrunculin are represented by black arrows.

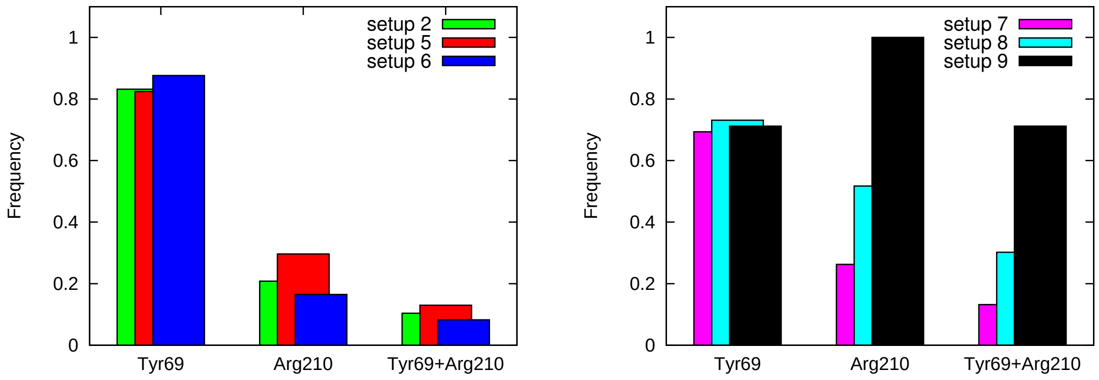
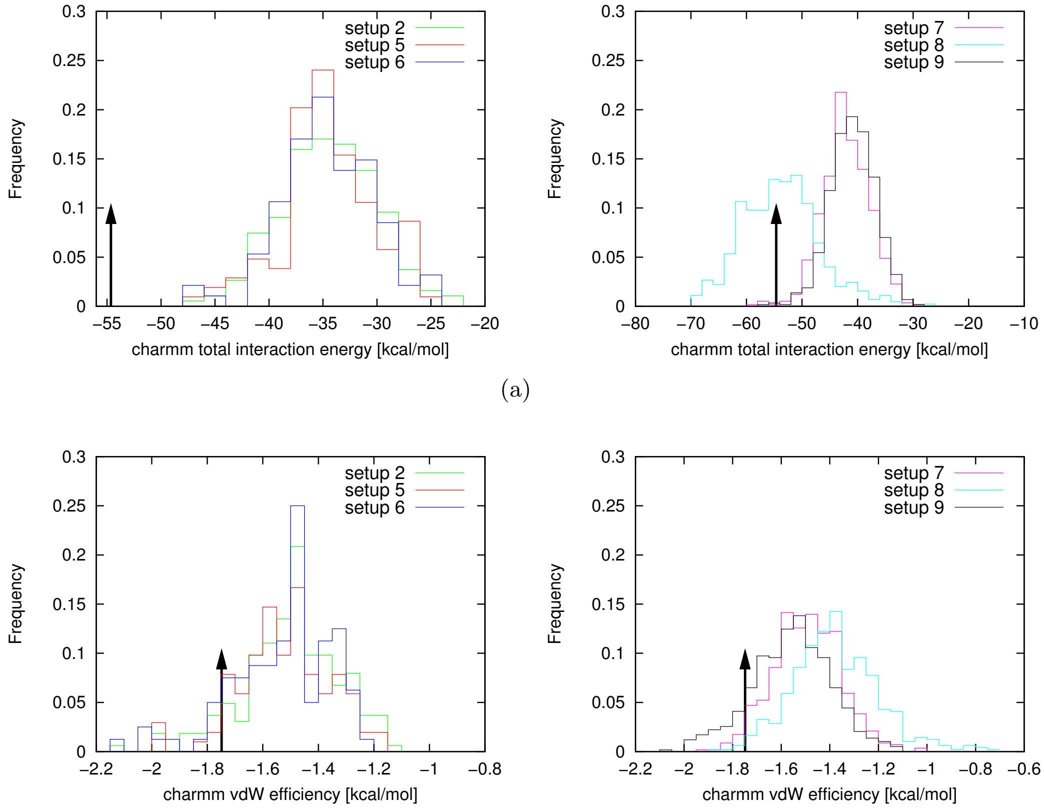


Figure 3.5: Percentage of generated molecules involved in hydrogen bonds to key residues Tyr69, Arg210, or both, for GANDI setups 2, 5, and 6 (left) and 7-9 (right).



(a)

(b)

Figure 3.6: Distribution of CHARMm protein-compound interaction energies for compounds generated using GANDI setups 2, 5, and 6 (left) and 7-9 (right). The black arrow shows the corresponding value for latrunculin. (a) CHARMm protein-compound total interaction energies. (b) CHARMm protein-compound van der Waals interaction efficiencies.

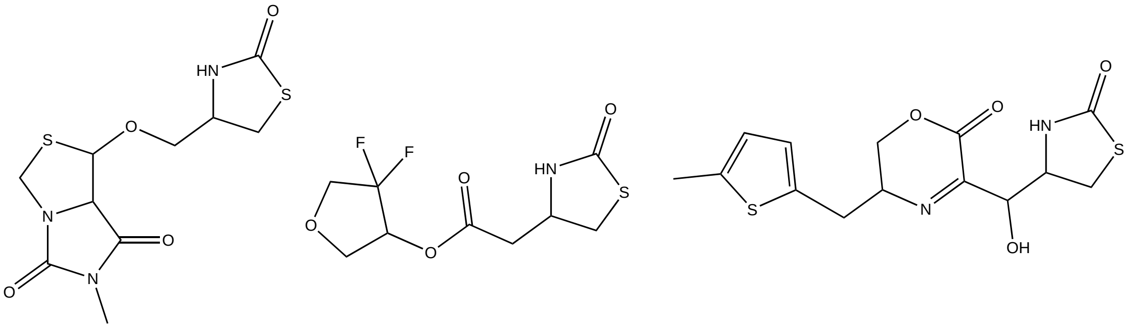


Figure 3.7: Examples of molecules generated by GANDI fulfilling both criteria of participation in hydrogen bonds to key residues and a CHARMm van der Waals interaction efficiency more favorable than -1.70 kcal/mol. The left and middle molecules were created using setup 2, the right molecule using setup 9. Note the presence of the hydroxyl group in the molecule generated by setup 9 which was included in the thiazolidinone anchor fragment.

Chapter 4

Inhibition of interdomain motion in G-actin by the natural product latrunculin: A molecular dynamics study

S. Rennebaum and A. Caflisch, *Proteins: Structure, Function, and Bioinformatics*, **2012**, doi: 10.1002/prot.24088

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Inhibition of interdomain motion in G-actin by the natural product latrunculin: A molecular dynamics study

Sandra Rennebaum and Amedeo Caflisch*

Department of Biochemistry, University of Zürich, Zürich 8057, Switzerland

ABSTRACT

As part of the cytoskeleton, actin is essential for the morphology, motility, and division of eukaryotic cells. Recent X-ray fiber diffraction studies have shown that the conformation of monomeric actin is flattened upon incorporation into the filament by a relative rotation of its two major domains. The antiproliferative activity of latrunculin, a macrolide toxin produced by sponges, seems to be related to its binding to monomeric actin and inhibition of polymerization. Yet, the mechanism of inhibition is not known in detail. Here, multiple explicit water molecular dynamics simulations show that latrunculin binding hinders the conformational transition related to actin polymerization. In particular, the presence of latrunculin at the interface of the two major domains of monomeric actin reduces the correlated displacement of Domain 2 with respect to Domain 1. Moreover, higher rotational flexibility between the two major domains is observed in the absence of ATP as compared to ATP-bound actin, offering a possible explanation as to why actin polymerizes more favorably in the absence of nucleotides.

Proteins 2012; 00:000–000.

Key words: actin; conformational transition; domain rotation; motional correlation; latrunculin.

INTRODUCTION

Actin plays an important role in various cellular processes such as cell motility, cell division, muscle contraction, and cytokinesis.[1] It is the most abundant protein in many eukaryotic cells and highly conserved among different organisms. The globular monomeric form, called G-actin, can reversibly assemble to form filamentous actin (F-actin) via a process controlled by a large number of actin-binding proteins. F-actin constitutes an integral part of the cytoskeleton. In tumor cells, the actin filament morphology is substantially altered which, together with its role in cell division, suggests that actin is a potential drug target.[2–5]

The first crystal structure of G-actin was published in 1990,[6] and today there are more than 80 actin crystal structures available in the Protein Data Bank[7] (PDB). Actin polymerization is inhibited by a number of natural products including latrunculins, a class of macrolide toxins produced by sponges in the Red Sea, including the genus *Latrunculia*, whence the name is derived.[8] The crystal structure of the G-actin/latrunculin A complex[9] shows the macrolide binding above the nucleotide binding site between the two major domains of G-actin
F1 (Fig. 1, top left), with its unique 2-thiazolidinone moiety
F2 buried deep in the cleft (Fig. 2, top). Latrunculins have antiproliferative, antiangiogenic, antimetastatic, and antimicrobial effects.[10–14] In addition, they reduce intraocular pressure in monkeys, such that they may be useful as treatment agents for glaucoma.[15] Aside from naturally occurring latrunculins, several synthetic analogs are likewise capable of disrupting actin filaments.[14,16–18]

Recent X-ray fiber diffraction data show the main conformational change in the G- to F-actin transition to be a relative rotation of the two major domains by about 20°, resulting in a flat actin monomer in the filament.[19] The structural flattening has been confirmed by cryo-EM[20,21] and a recently published F-actin model by Holmes and coworkers.[22]

Several computational studies have investigated the structural features and plasticity of G- and F-actin. Molecular dynamics (MD) simulations have shed light on the influence of the bound nucleotide on actin

---

Additional Supporting Information may be found in the online version of this article.
Grant sponsor: Swiss National Science Foundation (to A.C.). AQ1
*Correspondence to: Amedeo Caflisch, Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, Zürich 8057, Switzerland.
E-mail: caflisch@bioc.uzh.ch
Received 10 February 2012; Revised 22 March 2012; Accepted 30 March 2012

DOI: 10.1002/prot.24088

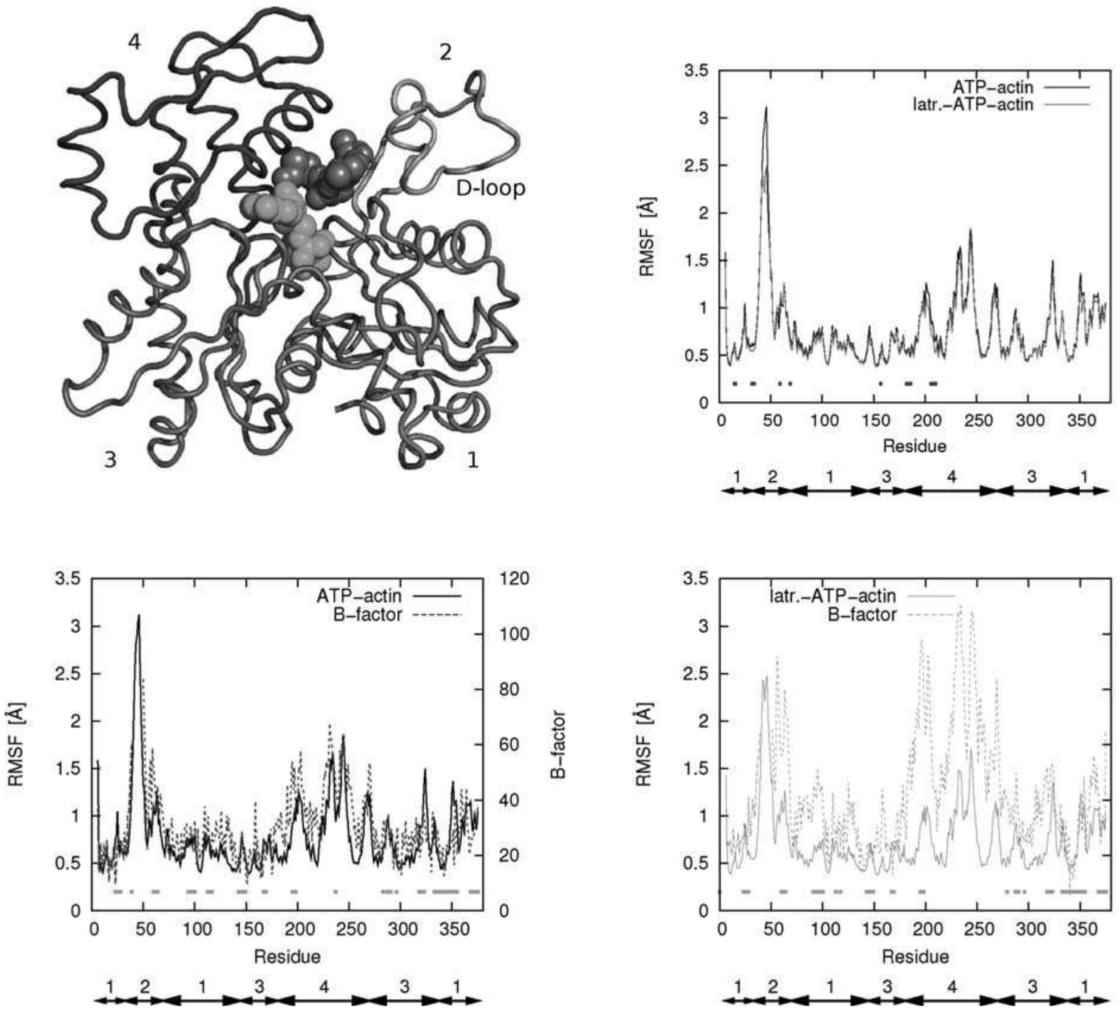


Figure 1

Structure and plasticity of G-actin. (Top, left) Structure of G-actin in complex with latrunculin A based on the 1ESV crystal structure. Actin is conventionally divided into two major domains. Domain 1 consists of Subdomains 1 (residues 1–32, 70–144, and 338–375, shown in red) and 2 (33–69, orange), while Domain 2 consists of Subdomains 3 (145–180 and 270–337, light blue) and 4 (181–269, dark blue). ATP (green) and latrunculin (gray) bind in a cleft between the two domains. (Top, right) RMSFs of C_α atoms in Å as a function of residue number for ATP–actin (black) and latr.–ATP–actin (red). The RMSF values are average values over simulation intervals of 5 ns, and the first 10 ns of each run were neglected. The subdomain numbering is given by the black arrows below the *x*-axis. The blue squares show the residues constituting the latrunculin binding site, that is, those residues with at least 50% of atoms within 5.0 Å of any atom of the inhibitor or residues forming a hydrogen bond to latrunculin in the 1ESV crystal structure (residues Gly15, Leu16, Pro32, Ile34, Gln59, Tyr69, Asp157, Gly182, Arg183, Thr186, Arg206, Glu207, and Arg210). (Bottom left and right) RMSF of C_α atoms in Å (black and red line, respectively, with *y*-axis on the left) and the crystallographic B-factors (blue and magenta dots, respectively, with *y*-axis on the right) for ATP–actin (bottom left) and latr.–ATP–actin (bottom right). Crystal contacts are shown by green squares for residues with one or more heavy atoms closer than 5 Å to heavy atoms of neighboring proteins in the crystal. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

conformation.[23–26] The conformation of the DNase I binding loop (D-loop) is still controversially discussed. Although Zheng *et al.* observed an unfolded D-loop in the ATP-bound and a folded loop in the ADP-bound state,[24] a nucleotide-dependence of the D-loop conformation could not be confirmed by Dalhaimer *et al.*[25] or Splettstoesser *et al.*[26] In a recent metadynamics simulation study, it was found that the folded and unfolded states of the D-loop are similarly stable in ADP–actin.[23] Moreover, MD simulations indicate that the predominant form of G-actin is the closed and twisted conformation, independent of the nature of the bound nucleotide.[25] The open conformation of actin, observed only in complex with profilin, was found to be unstable upon

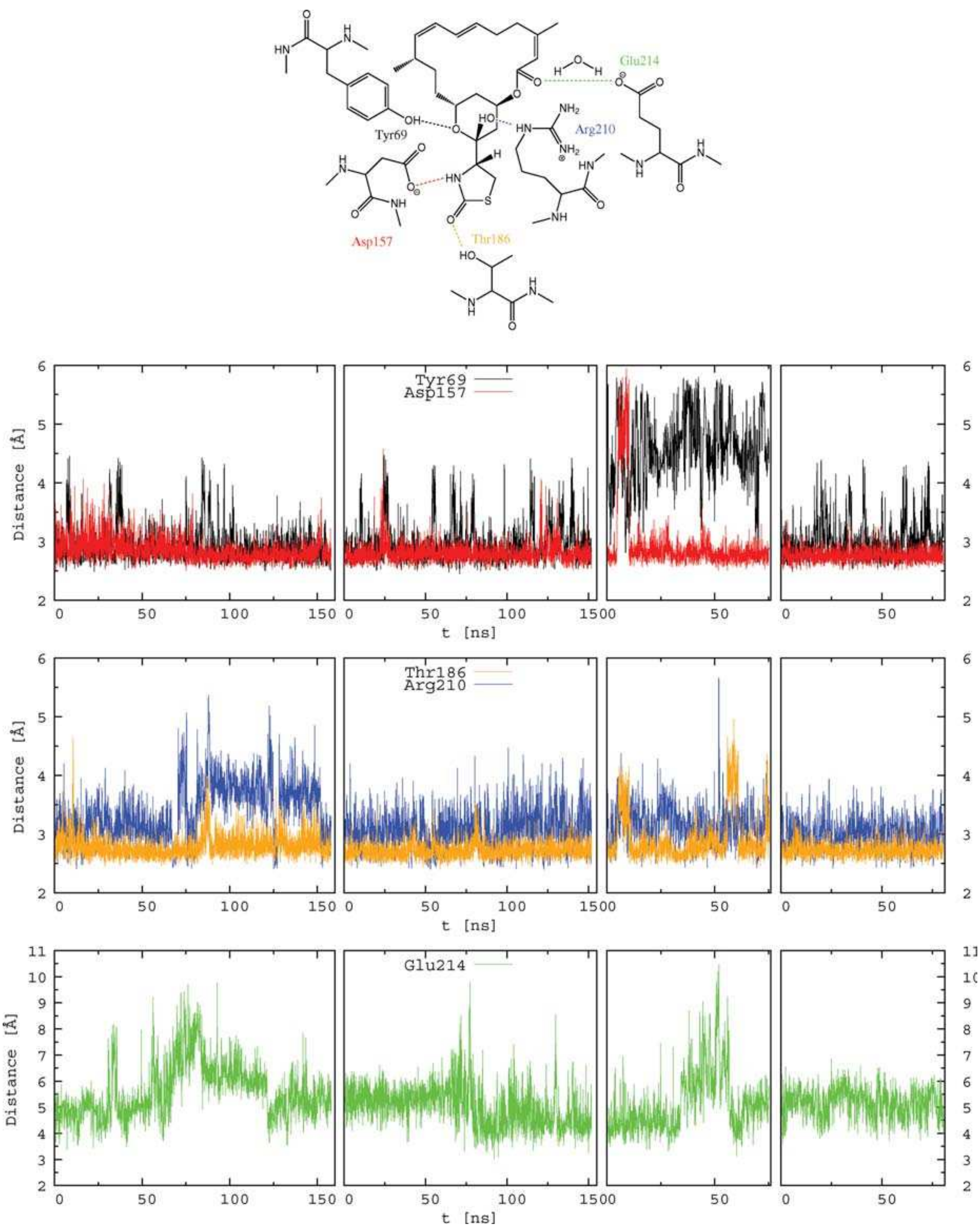


Figure 2

Stability of latrunculin binding mode along MD simulations. (Top) Two-dimensional scheme of latrunculin A and G-actin residues involved in hydrogen bonds. (Middle and bottom) Time series of hydrogen bond distances, that is, the distance between donor and acceptor atoms. The colors are consistent with those used in the top panel. The individual MD runs are separated by black vertical lines. Note the different *y*-axis range for the hydrogen bond to Glu214, which is water mediated.

removal of profilin.[26,27] Recent MD studies of the Oda model[19] and the latest Holmes model[22] of F-actin indicate that the interdomain twist of F-actin increases slightly during the simulations.[22,28] However, the twist angle still remains significantly smaller than in G-actin simulations for both models. In recent MD simulations,[26] a "superclosed" G-actin conformation was observed, which resembles the structure in F-actin models.[19–22] Furthermore, water molecules in the nucleotide binding site have been shown to influence actin conformation in MD simulations.[29] The influence of the natural product phalloidin on the actin filament has been investigated by explicit solvent MD simulations.[28] Interestingly, during these simulations a displacement was observed from the original position proposed by the experimentalists toward a site with more adjacent interstrand contacts between subunits along the short-pitch helix, which is congruent with the filament stiffening effect of phalloidin.[28]

As of today, no computational study of the effect of latrunculin on the dynamical properties of G-actin has been reported. Here, we investigate the motional correlation between subdomains in monomeric actin and the influence of latrunculin on the relative rotation of the two major domains by explicit solvent MD simulations. Three systems are investigated in detail: apo actin, ATP-bound actin, and ATP-bound actin in complex with latrunculin A. Multiple MD runs for each system are carried out for a total simulation time of about 1.2 μs. Simulations of nucleotide-free actin indicate a higher interdomain rotational flexibility compared to the ATP-bound state, which is congruent with the observation that nucleotide-free actin polymerizes more favorably than ATP–actin.[30] Moreover, the simulations in the presence of latrunculin show that binding of this natural product prevents the relative rotation of the two major domains, which is necessary for the G- to F-transition, thus interfering with actin polymerization.

MATERIALS AND METHODS

Preparation of the structures

The coordinates of the inhibitor-free[31] and -bound[9] actin were downloaded from the PDB database (PDB ID: 1EQY and 1ESV, respectively). As residues 1–5 are missing in the X-ray structures, the $—COCH_3$ group was added to the N-terminal Thr6. The C-termini were considered negatively charged.

Due to the high flexibility of the D-loop, the coordinates of its residues are not present in the PDB files (residues 40–49 in 1EQY and 40–50 in 1ESV). As in previously published MD studies of G-actin,[23–25,29] initial D-loop coordinates were obtained from an actin crystal structure with the D-loop residues resolved. Using SWISS-MODEL,[32] the PDB structure 3DAW was selected as template for the two following reasons. First, the 1EQY and 1ESV coordinate sets are ATP-bound actin structures, which is also the case for 3DAW. Second, in 3DAW, the protein used for cocrystallization (which is a domain of twinfilin) binds between the actin Subdomains 1 and 3, which are located far away from the D-loop, and thus the presence of the twinfilin domain does not directly influence the conformation of the D-loop. Similar to the procedure described in Ref. 24, after fitting the crystal structures (1EQY and 1ESV) to the template using the C_α atoms present in all structures, the missing coordinates of the D-loop were taken from the template and inserted into the 1EQY and 1ESV structure. To relax elongated bonds after the coordinate transfer, an energy minimization of 100 steps of the steepest descent and 1000 steps of the adopted basis Newton–Raphson algorithm was carried out using the program CHARMM[33,34] and the CHARMM22 force field.[35] During the minimization, all coordinates present in the X-ray structures were kept fixed. Note that in this study, the coordinates of the D-loop are not used for structural alignment or calculation of root mean square deviation (RMSD). As there is no crystal structure of the ATP-free actin available, the coordinates of ATP and the associated calcium ion were removed from the 1EQY structure. The resulting structure is referred to as "apo actin."

MD simulations

To reproduce neutral pH conditions, the side chains of aspartates and glutamates were negatively charged, those of lysines and arginines were positively charged, and histidines were considered neutral. The protein was immersed in an orthorhombic box of pre-equilibrated water molecules. The size of the box was chosen to have a minimal distance of 13 Å between the boundary and any atom of the protein. VMD[36] was used for setting up the simulation system, while minimization, heating, and production runs were performed with NAMD[37] using AQ2
the CHARMM22 force field[35] and the TIP3P model of water. For the parameters of latrunculin A, the CGenFF force field[38] was used. Periodic boundary conditions were applied, and the particle-mesh Ewald approach[39] was used for the long-range electrostatics. The van der Waals interactions were truncated at a cutoff of 12 Å, and a switch function was applied starting at 10 Å. The MD simulations were carried out at constant temperature (298 K) and constant pressure (1 atm) with a time step of 2 fs using the SHAKE algorithm[40] to fix the length of covalent bonds involving hydrogen atoms.

Twist angle

Twist angles between Domains 1 and 2 were calculated as described in Ref. 41, where the twist angle is defined as the angle between two planes, one containing the C_α

atom of residue Gly55 (Subdomain 2) and the axis of rotation, the other containing Glu207 (Subdomain 4) and the axis of rotation. Each MD snapshot was superimposed to the monomeric ADP–actin structure (PDB ID 1J6Z).[41] The axis of rotation was determined between 1J6Z and the F-actin subunit (PDB ID 2ZWH) using DynDom.[42] In addition to the twist angle, time series of the dihedral angle between the centers of mass of the four subdomains (also called "propeller angle"[26]) were computed. The propeller angle time series are very similar to the twist angle time series and therefore not shown.

Motional correlations from MD trajectories

Normalized fluctuation correlations between pairs of C_α atoms were calculated using the DCC[43] algorithm as implemented in WORDOM [see Eq. (10) in Ref. 44]. Their values range from -1 (for a fully anticorrelated motion between two C_α atoms, i.e., motion in opposite direction) through 0 (indicating no correlation) to $+1$ (for a fully correlated motion).

The linear mutual information (LMI)[45,46] algorithm was utilized as a second method to compute motional correlations [see Eq. (11) in Ref. 44. LMI values vary from 0 indicating lack of any correlation to $+1$, which is complete correlation between atomic displacements. Anticorrelation is not captured by the mutual information measure. Correlations between perpendicular motions are estimated by the LMI but not the DCC method.

It should be noted that the motional correlations depend strongly on the choice of atoms and reference frame for alignment prior to the correlation analysis. The C_α atoms of Domain 1, excluding residues 40–50, were utilized for the alignment to measure the degree of intradomain correlation of Domain 2. The average structure calculated from the MD trajectories was used as reference frame for each system. To test the robustness of the choice of reference frame, the first 20 ns of each trajectory were removed, and the average structure recalculated. Using the resulting structure as reference, frame led to essentially identical covariance matrices.

Accession numbers

The structures used in this study were obtained from the PDB database under accession codes 1EQY, 1ESV, 1J6Z, 2ZWH, and 3DAW.

RESULTS AND DISCUSSION

ATP-bound G-actin is considered as reference and most of the analysis focuses on the differences with respect to this system referred to as "ATP-actin." The tripartite complex is called "latr.–ATP–actin" and the nucleotide-free G-actin is called "apo actin." When "F-actin" is mentioned, the Oda model[19] is referred to. Table I gives an overview of the performed MD simulations. T1

Table I
Summary of Performed MD Simulations

System	Starting str. (PDB ID)	Latrunculin presence	ATP presence	No. of runs[a]	Length (ns)[b]
ATP–actin	1EQY	No	Yes	4	152, 136, 91, 89
Latr.–ATP–actin	1ESV	Yes	Yes	4	158, 152, 75, 80
Apo actin[c]	1EQY	No	No	2	80, 77

[a] Multiple runs were started using different seeds to generate a random distribution of the initial velocities.
[b] The different lengths of the individual runs are due to manual stopping after the first 150 ns or 75 ns.
[c] Since no crystal structure of ATP-free actin is available, the coordinates of ATP and the associated calcium ion were removed from the 1EQY structure.

Overall stability and flexibility

The low values of the C_α RMSD from the X-ray structure of G-actin used as starting conformation indicate that the overall structural stability is preserved in all MD runs (Supporting Information, Fig. S1). Although apo actin denatures in solution,[47] its stability is preserved for the entire duration of the MD simulations. Apo actin was also found to be stable in a previously published MD simulation of 8 ns.[25] As the rate constant of the denaturation is 0.2 s^{-1},[47] unfolding is not expected to occur even in the 100 ns time scale of the present MD simulations. There is a good correlation between the root mean square fluctuation (RMSF) of the C_α atoms along the MD simulations and the crystallographic temperature factors for ATP–actin (Fig. 1). The low RMSD values and reasonable fluctuations indicate that the force field and simulation protocol are adequate for investigating the dynamical properties of G-actin. The nearly perfect overlap of the calculated RMSF of ATP–actin and latr.–ATP–actin (Fig. 1, top right) suggests that latrunculin influences marginally the fluctuations of the actin backbone on the nanosecond time scale. Thus, the higher *B*-factors for the latr.–ATP–actin complex than in the absence of latrunculin, particularly for Subdomain 4, might originate from disorder in the crystal of the former. In all MD runs, the highest mobility is observed for the D-loop (residues 40–50) in Subdomain 2, which is the most flexible part of G-actin, and residues 220–250 in Subdomain 4, in agreement with recently published MD results of the ADP-bound state.[22] The absence of ATP does not have a strong effect on the C_α RMSF either (Supporting Information, Fig. S2).

It is interesting to analyze the stability of the latrunculin binding mode. Latrunculin stays in its binding site for the entire duration of the MD simulations as shown by the RMSD time series of the heavy atoms of latrunculin that oscillate between 1 and 2.5 Å during most of the runs (Supporting Information, Fig. S3). Moreover, the two intermolecular hydrogen bonds between the thiazolidinone ring of latrunculin and the side chains of Asp157 and Thr186 are stable (Fig. 2). A slightly lower stability is observed for the hydrogen bonds between the tetrahydropyran moiety and the side chains of Tyr69 and Arg210, while the water-bridged hydrogen bond between the lactone carbonyl and the side chain of Glu214 shows relatively strong fluctuations.

Evidence for the hindrance of interdomain rotation by latrunculin

F3

The relative motion between domains can be measured by the RMSD from G-actin and F-actin after overlap of the C_α atoms of Domain 1. In contrast to ATP–actin, only a negligible number of MD snapshots have a RMSD from F-actin below 4 Å in the latrunculin-bound form (Fig. 3, middle). The contrast is even more pronounced between latr.–ATP–actin and apo actin. The percentage of MD snapshots with a RMSD from F-actin below 3.5 Å is 4.7, 4.4, and 0% for apo actin, ATP–actin, and latr.–ATP–actin, respectively. Using a threshold of 4 Å the corresponding values are 15.1, 11.6, and 0.3%. These simulation results suggest that latrunculin binding decreases the probability of the protein adopting an F-actin-like conformation, whereas the absence of ATP increases the probability. The anticorrelation between the RMSD from G- and F-actin suggests that a conformational change increasing the deviation from G-actin augments also the structural similarity to the conformation observed in the filament. According to Figure 3, the degree of anticorrelation for the three investigated systems is the lowest for latr.–ATP–actin and the highest for apo actin. Thus, the presence of latrunculin interferes with the structural rearrangements required for polymerization.

F4

Another interesting measure of the relative motion of the two domains in the MD trajectories is the fluctuation correlation between pairs of residues, which was computed with two different methods (Fig. 4). In contrast to the dynamic cross correlation (DCC) method, the LMI algorithm is able to estimate correlations between perpendicular motions (see "Materials and Methods" section). For ATP–actin a high degree of correlated movement is observed within the major Domain 2, which indicates that Domain 2 moves almost as a rigid body with respect to Domain 1. While the fluctuations of some segments, for example, residues 157–172 and 275–292, appear uncorrelated to the movement of a large part of Domain 2 according to the DCC results, using the LMI method reveals a high degree of correlation and suggests a perpendicular movement of these residues with respect to a large number of residues in Domain 2. The matrix of difference of motional correlation (Fig. 4, bottom) shows a reduced intradomain correlation for Domain 2 in the presence of latrunculin, indicating that Domain 2 moves to a lesser extent as a rigid body relative to Domain 1 in the latrunculin complex.

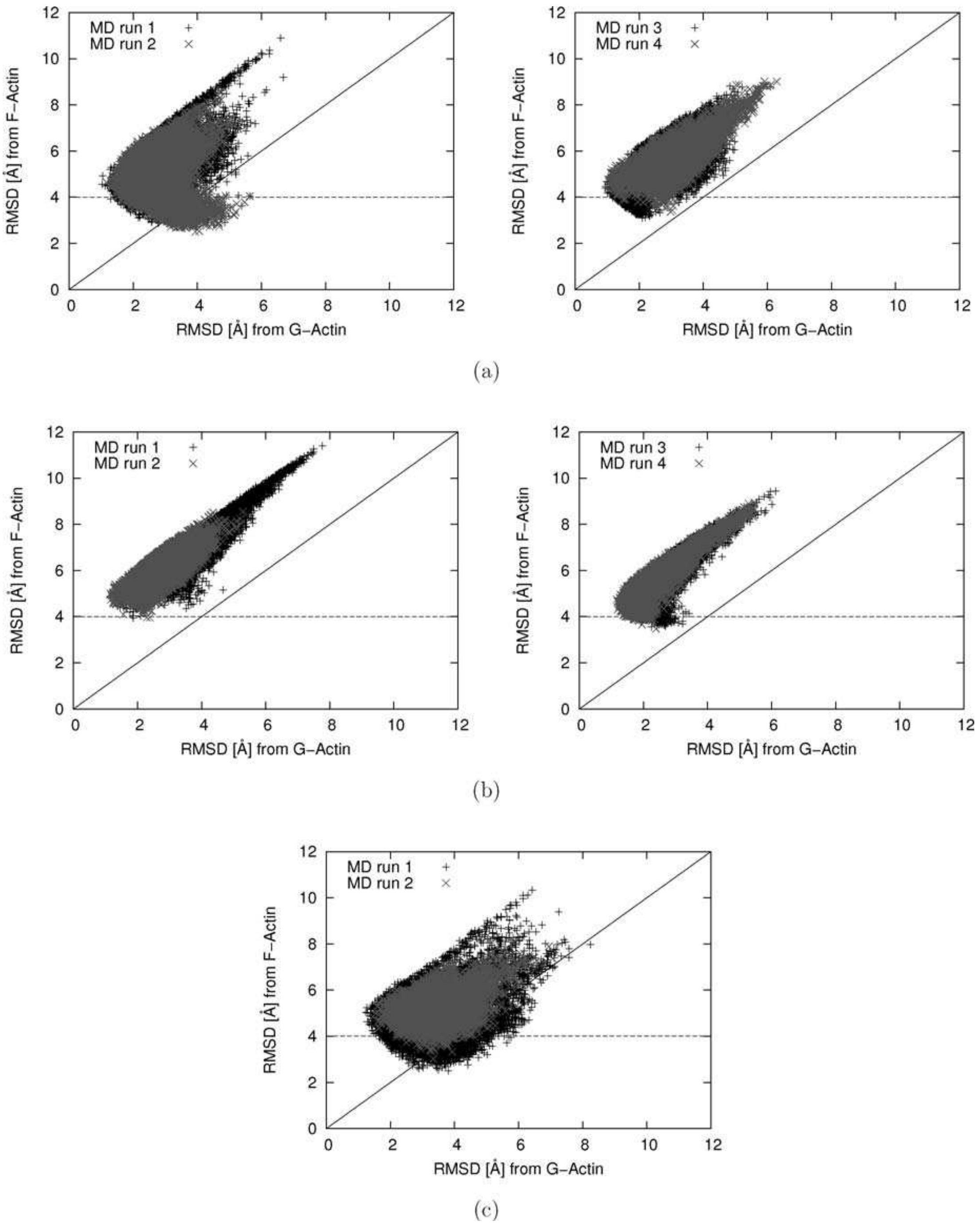
F5

The relative rotation of the two main domains can also be monitored by the twist angle (Fig. 5), which is defined as the angle between two planes intersecting at the axis of rotation, with one plane containing the C_α atom of residue Gly55 (Subdomain 2) and the other containing the C_α atom of Glu207 (Subdomain 4).[41] The twist angle in F-actin is 8°, and it ranges between 16 and 25° in G-actin (depending on the crystal structure) with an average of 20°. Overall, higher twist angles are observed for latr.–ATP–actin than for ATP–actin. In the simulations of the latter, there are several events where the twist angle is almost reduced to the value in F-actin or even further, resulting in a flattened structure similar to the conformation in the filament. As an example, in the time interval 75–97 ns of the second run of the ATP–actin simulations the twist angle is close to the one of F-actin (Fig. 5(b)]. Note also that during this time interval, the RMSD from F-actin and G-actin after superposition of the C_α atoms of Domain 1 (Supporting Information, Fig. S4) are approximately equal. A flat conformation of ATP-bound G-actin has also been observed in 4 out of 20 explicit water MD simulations in a previous study (where it was called a "superclosed" state).[26] The authors suggested that the superclosed state is not the predominant form of ATP–actin in equilibrium, which is a plausible reason why this state has not been observed crystallographically. In contrast to the flattening observed in the absence of latrunculin, over the entire course of the four MD runs of latr.–ATP–actin, the twist angle remained close to or even higher than the one of G-actin. Thus, latrunculin binding to monomeric actin prevents the relative rotation of the two major domains required for the polymerization process. Interestingly, there are simulation segments during which an increase in RMSD from both G- and F-actin relative to the simulation average (Supporting Information, Fig. S1) correlates with an increase in twist angle (e.g., the time intervals 80–90 ns in the first run of ATP–actin, 3–10 ns and 17–24 ns in the fourth run of latr.–ATP–actin, and 33–39 ns in the first run of apo actin). Here, the two major domains of actin rotate in the opposite direction compared to the flattening in the G- to F-actin transition.

Finally, the interdomain rotation occurs most frequently in simulations of apo actin. There are more flattening events in the simulations of the nucleotide-free structure than in those of ATP–actin, though the former sampling is only about one-third of the latter. Moreover, twist angle values of about 5° are reached only in the absence of ATP. These findings suggest that ATP slightly hinders the relative displacements of the two main domains of G-actin.

CONCLUSIONS

Explicit solvent MD simulations of monomeric actin in the presence and absence of latrunculin have been carried out to study the influence of the binding of this natural



(a)

(b)

(c)

Figure 3

Deviation of MD snapshots from G-actin and F-actin. Scatter plots of RMSD from G- and F-actin, calculated for C_α atoms of Domain 2 upon overlap of C_α atoms of Domain 1 (excluding residues 40–50 of the flexible D-loop), using the structure 2ZWH as reference for F-actin. (**a**) ATP–actin, (**b**) latr.–ATP–actin, and (**c**) apo actin. The reference structure for G-actin is 1EQY in (a) and (c) and 1ESV in (b). For better visibility, MD Runs 3 and 4 are shown separately from Runs 1 and 2 in (a) and (b). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

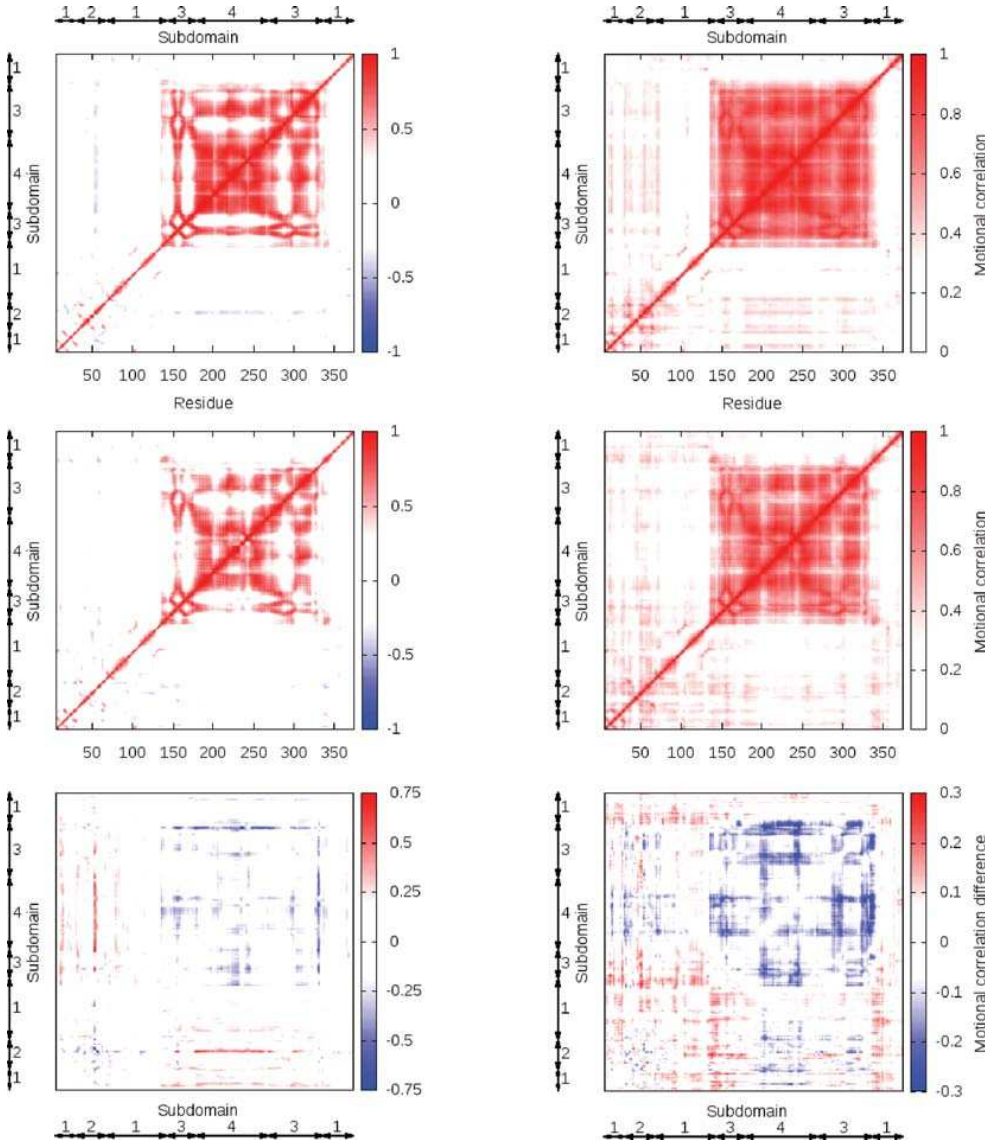
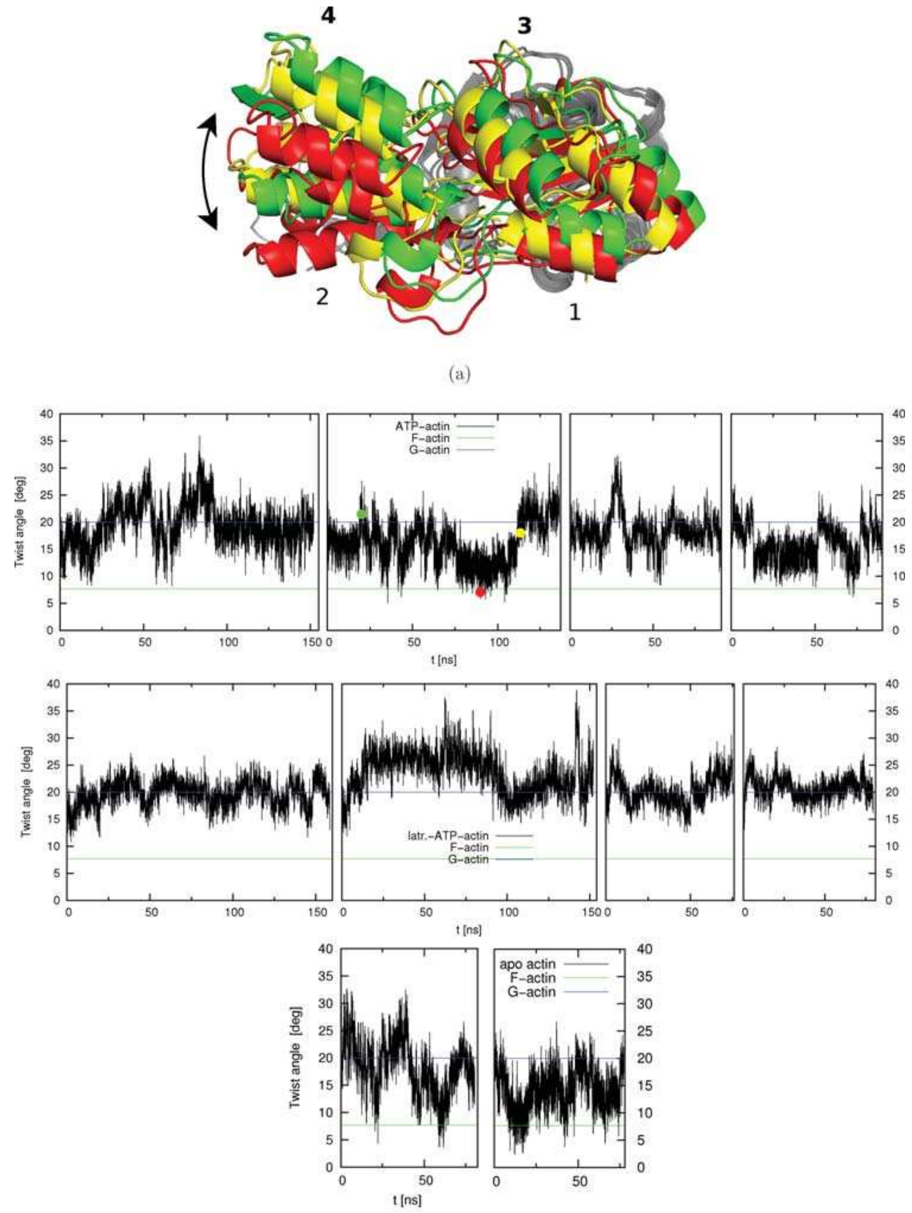


Figure 4

Latrunculin reduces correlated displacement of Domain 2 versus Domain 1. Motional correlations from MD trajectories of ATP–actin (top) and latr.–ATP–actin (middle) calculated by the DCC (left panels) and LMI (right panels) algorithm. The C$_{\alpha}$ atoms of Domain 1 (excluding the D-loop) were superimposed prior to the correlation analysis. The color scale ranges from blue (anticorrelation) to white (no correlation) to red (correlation) for the DCC algorithm and from white (no correlation) to red (correlation) for the LMI method. The subdomain numbering is given by the black arrows. Note that the two algorithms give similar qualitative results and in particular weaker correlation within the major Domain 2, that is, Subdomains 3 and 4, in the presence of latrunculin. The differences between the two algorithms, and in particular the higher correlation reported by LMI than DCC, originate from the fact that only LMI takes into account correlated perpendicular motion.[44] (Bottom) Difference between covariance matrices of latr.–ATP–actin and ATP–actin.

product on the plasticity of G-actin. The simulation results provide evidence that latrunculin prevents actin polymerization by hindering the rotation of the two major domains associated with the G- to F-actin transition. Time series of the twist angle show no substantial flattening for latr.–ATP–actin in contrast to simulations of ATP–actin and apo actin. Moreover, cross correlations of atomic displacements indicate a lower degree of rigid body movement of the two domains relative to each other upon binding of latrunculin. The rotational flexibility of the two domains in G-actin increases in the following order for the three investigated systems: latr–ATP–actin $\ll$ ATP–actin $<$ apo actin.



(a)

(b)

Figure 5

Twist angle analysis shows that latrunculin prevents interdomain motion. (**a**) Side view on Subdomains 3 and 4 of three selected snapshots (green, red, and yellow) from the ATP–actin runs after superposition of C_α atoms of Subdomains 1 and 2 (colored in gray) excluding the D-loop. The arrow illustrates the relative rotation of Subdomain 4. The axis of rotation passes through Subdomains 1 and 3 (not shown). (**b**) Twist angle time series of ATP–actin (top), latr.–ATP–actin (middle), and apo actin (bottom). The three colored circles in the ATP–actin time series correspond to the three snapshots shown in (a). The blue horizontal line shows the mean value of the twist angle measured on 83 crystal structure of G-actin, whereas the green horizontal line indicates the twist angle in the F-actin structure of the Oda model.[19] The individual MD runs are separated by black vertical lines.

This conclusion is consistent with the experimental observations that nucleotide-free actin polymerizes more easily than ATP-bound actin,[30] whereas latrunculin-bound actin is not able to polymerize.[9]

Considering that both ATP and latrunculin bind in a cleft between the two major domains of monomeric actin, the MD results seem plausible: the domains are able to rotate more freely, when their relative movement is not restricted by the binding of ligands. The inhibitory mechanism of latrunculin proposed here on the basis of the MD simulations could be verified experimentally by mutating one or more of the actin side chains in contact with latrunculin into bulkier ones, for example, Leu16Trp, Ile34Trp, and/or Tyr69Trp. Finally, motivated by the MD simulation results, we are currently carrying out *de novo* design[48] of small molecules that bind in the same cleft as latrunculin as potential anti-cancer compounds.

ACKNOWLEDGMENTS

The authors thank Emilie Frugier and Tim Knehans for carefully reading the manuscript. They thank the anonymous reviewers for interesting suggestions. The calculations were carried out on the Schrödinger computer cluster of the University of Zürich (Switzerland).

REFERENCES

1. Schmidt A, Hall MN. Signaling to the actin cytoskeleton. Annu Rev Cell Dev Biol 1998;14:305–338.
2. Jordan MA, Wilson L. Microtubules and actin filaments: dynamic targets for cancer chemotherapy. Curr Opin Cell Biol 1998;10: 123–130.
3. Janmey PA, Chaponnier C. Medical aspects of the actin cytoskeleton. Curr Opin Cell Biol 1995;7:111–117.
4. Giganti A, Friederich E. The actin cytoskeleton as a therapeutic target: state of the art and future directions. Prog Cell Cycle Res 2003;5:511–525.
5. Lambrechts A, Van Troys M, Ampe C. The actin cytoskeleton in normal and pathological cell motility. Int J Biochem Cell Biol 2004;36:1890–1909.
6. Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC. Atomic structure of the actin: DNase I complex. Nature 1990;347:37–44.
7. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
8. Nèeman I, Fishelson L, Kashman Y. Isolation of a new toxin from the sponge *Latrunculia magnifica* in the Gulf of Aquaba (Red Sea). Mar Biol 1975;30:293–296.
9. Morton WM, Ayscough KR, McLaughlin PJ. Latrunculin alters the actin-monomer subunit interface to prevent polymerization. Nat Cell Biol 2000;2:376–378.
10. Spector I, Braet F, Shochet NR, Bubb MR. New anti-actin drugs in the study of the organization and function of the actin cytoskeleton. Microsc Res Tech 1999;47:18–37.
11. Longley RE, McConnell OJ, Essich E, Harmody D. Evaluation of marine sponge metabolites for cytotoxicity and signal transduction activity. J Nat Prod 1993;56:915–920.
12. Konishi H, Kikuchi S, Ochiai T, Ikoma H, Kubota T, Ichikawa D, Fujiwara H, Okamoto K, Sakakura C, Sonoyama T, Kokuba Y, Sasaki H, Matsui T, Otsuji E. Latrunculin A has a strong anticancer effect in a peritoneal dissemination model of human gastric cancer in mice. Anticancer Res 2009;29:2091–2097.
13. El Sayed KA, Khanfar MA, Shallal HM, Muralidharan A, Awate B, Youssef DTA, Liu Y, Zhou YD, Nagle DG, Shah G. Latrunculin A and its C-17-O-carbamates inhibit prostate tumor cell invasion and HIF-1 activation in breast tumor cells. J Nat Prod 2008;71:396–402.
14. Khanfar MA, Youssef DTA, El Sayed KA. Semisynthetic latrunculin derivatives as inhibitors of metastatic breast cancer: biological evaluations, preliminary structure–activity relationship and molecular modeling studies. ChemMedChem 2010;5:274–285.
15. Peterson JA, Tian B, Geiger B, Kaufman PL. Effect of latrunculin B on outflow facility in monkeys. Exp Eye Res 2000;70:307–313.
16. Fürstner A, Kirk D, Fenster MD, Aïssa C, De Souza D, Müller O. Diverted total synthesis: preparation of a focused library of latrunculin analogues and evaluation of their actin-binding properties. Proc Natl Acad Sci USA 2005;102:8103–8108.
17. Fürstner A, Kirk D, Fenster MD, Aïssa C, De Souza D, Nevado C, Tuttle T, Thiel W, Müller O. Latrunculin analogues with improved biological profiles by "Diverted otal Synthesis": preparation, evaluation, and computational analysis. Chem Eur J 2007;13:135–149.
18. Amagata T, Johnson TA, Cichewicz RH, Tenney K, Mooberry SL, Media J, Edelstein M, Valeriote FA, Crews P. Interrogating the bioactive pharmacophore of the latrunculin chemotype by investigating the metabolites of two taxonomically unrelated sponges. J Med Chem 2008;51:7234–7242.
19. Oda T, Iwasa M, Aihara T, Maéda Y, Narita A. The nature of the globular- to fibrous-actin transition. Nature 2009;457:441–445.
20. Fujii T, Iwane AH, Yanagida T, Namba K. Direct visualization of secondary structures of F-actin by electron cryomicroscopy. Nature 2010;467:724–728.
21. Murakami K, Yasunaga T, Noguchi TQ, Gomibuchi Y, Ngo KX, Uyeda TQ, Wakabayashi T. Structural basis for actin assembly, activation of ATP hydrolysis, and delayed phosphate release. Cell 2010;143:275–287.
22. Splettstoesser T, Holmes KC, Noé F, Smith JC. Structural modeling and molecular dynamics simulation of the actin filament. Proteins: Struct Funct Bioinform 2011;79:2033–2043.
23. Pfaendtner J, Branduardi D, Parrinello M, Pollard TD, Voth GA. Nucleotide-dependent conformational states of actin. Proc Natl Acad Sci USA 2009;106:12723–12728.
24. Zheng X, Diraviyam K, Sept D. Nucleotide effects on the structure and dynamics of actin. Biophys J 2007;93:1277–1283.
25. Dalhaimer P, Pollard TD, Nolen BJ. Nucleotide-mediated conformational changes of monomeric actin and Arp3 studied by molecular dynamics simulations. J Mol Biol 2008;376:166–183.
26. Splettstoesser T, Noé F, Oda T, Smith JC. Nucleotide-dependence of G-actin conformation from multiple molecular dynamics simulations and observation of a putatively polymerization-competent superclosed state. Proteins: Struct Funct Bioinform 2009;76: 353–364.
27. Minehardt TJ, Kollman PA, Cooke R, Pate E. The open nucleotide pocket of the profilin/actin X-ray structure is unstable and closes in the absence of profilin. Biophys J 2006;90:2445–2449.
28. Pfaendtner J, Lyman E, Pollard TD, Voth GA. Structure and dynamics of the actin filament. J Mol Biol 2010;396:252–263.
29. Saunders MG, Voth GA. Water molecules in the nucleotide binding cleft of actin: effects on subunit conformation and implications for ATP hydrolysis. J Mol Biol 2011;413:279–291.
30. De La Cruz EM, Mandinova A, Steinmetz MO, Stoffler D, Aebi U, Pollard TD. Polymerization and structure of nucleotide-free actin filaments. J Mol Biol 2000;295:517–526.
31. McLaughlin PJ, Gooch JT, Mannherz HG, Weeds AG. Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. Nature 1993;364:685–692.
32. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 2006;22:195–201.
33. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4:187–217.
34. Brooks BR, Brooks CL, MacKerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW,

Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545–1614.

35. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
36. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38.
37. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. Scalable molecular dynamics with NAMD. J Comput Chem 2005;26:1781–1802.
38. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, MacKerell AD. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J Comput Chem 2010;31:671–690.
39. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. J Chem Phys 1995;103:8577–8593.
40. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equation of motion of a system with constraints: molecular dynamics of *n*-alkanes. J Comput Phys 1977;23:327–341.
41. Oda T, Maéda Y. Multiple conformations of F-actin. Structure 2010;18:761–767.
42. Hayward S, Berendsen HJC. Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme. Proteins: Struct Funct Genet 1998;30:144–154.
43. McCammon JA, Harvey S. Dynamics of proteins and nucleic acids. Cambridge: Cambridge University Press; 1987.
44. Seeber M, Felline A, Raimondi F, Muff S, Friedman R, Rao F, Caflisch A, Fanelli F. Wordom: a user-friendly program for the analysis of molecular structures, trajectories, and free energy surfaces. J Comput Chem 2011;32:1183–1194.
45. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E 2004;69:066138–066154.
46. Lange OF, Grubmüller H. Generalized correlation for biomolecular dynamics. Proteins: Struct Funct Bioinform 2006;62:1053–1061.
47. Kinosian HJ, Selden LA, Gershman LC, Estes JE. Non-muscle actin filament elongation from complexes of profilin with nucleotide-free actin and divalent cation-free ATP–actin. Biochemistry 2004;43:6253–6260.
48. Dey F, Caflisch A. Fragment-based de novo ligand design by multiobjective evolutionary optimization. J Chem Inf Model 2008;48:679–690.

Conclusions

Actin, a protein that plays a pivotal role in the functioning of eukaryotic cells, was the focus of the present thesis. The only compounds known to bind to actin are structurally complex natural products or analogues thereof, which are difficult to synthesize or extract from natural sources.

To identify new lead structures for the design of novel actin inhibitors, a virtual screening campaign of natural product-derived compounds was carried out targeting the binding site at the so-called pointed end of actin. The binding behavior of top-ranking compounds selected by consensus scoring was investigated by molecular dynamics (MD) simulations. Those molecules consistently found to unbind within a few nanoseconds in multiple independent MD runs were discarded as leads, in accordance with the assumption that low unbinding times correlate with poor binding affinities. The applied approach is useful to eliminate molecules with a low chance to bind at an early stage of a drug discovery program. The lack of accuracy of scoring functions remains one of the principle limitations of virtual high-throughput docking; even the use of multiple scoring functions, as in this study, does not guarantee that only high-affinity binders are selected. MD simulations allow molecules predicted to have high binding affinities by docking to be discarded if their binding modes are unstable in MD. This way, time-consuming experiments are avoided for molecules having poor binding affinities.

A single exponential behavior was observed for the kinetics of ligand unbinding, indicating the existence of one predominant free energy barrier, the barrier of unbinding. The free energy barriers between different binding modes are much lower [1]. A small number of compounds were identified with relatively stable binding behavior and high unbinding times, suggesting a binding affinity within the micromolar range and their potential as lead compounds in the design of novel actin inhibitors.

The binding behavior of one of the top-ranking compounds selected by consensus scoring was optimized by performing chemical modifications according to the most populated structural clusters in the MD simulations. Addition of only two heavy atoms resulted in a significant increase in the stability of the most populated binding mode, indicating a higher binding affinity of the optimized relative to the original compound.

In chapter 3, putative actin inhibitors were identified using *de novo* design, another method for computer-aided drug design where molecules are built “from scratch”. The thiazolidinone moiety of latrunculin was used as an anchor from which additional fragments were grown, and compounds forming hydrogen bonds analogous to latrunculin and more favorable van der Waals interaction efficiencies were identified.

In vitro assays of the compounds obtained from the virtual screening and the *de novo* design campaigns are currently being performed at the Department of Organic Chemistry at the University of Zürich to determine their binding affinity and ability to inhibit actin polymerization.

In chapter 4, MD simulations were carried out to investigate a basic research question: the mechanism by which the natural product latrunculin inhibits actin polymerization. As recently shown by X-ray fiber diffraction [2], the propeller-like twist of monomeric actin is flattened in the G- to F-actin transition by an interdomain rotation of appr. 20 degrees. The MD study presented in chapter 4 provides evidence that latrunculin prevents this conformational change, thereby inhibiting actin polymerization. The presence of latrunculin reduces the extent of rigid body movement of the two major domains relative to one another as shown by normalized fluctuation correlations. Furthermore, measuring the degree of twist between the two actin domains over the course of the MD simulations showed no substantial flattening of actin in the presence of latrunculin. In contrast, in absence of latrunculin, the interdomain rotation was observed for ATP-bound actin and even more frequently for ATP-free actin. The simulation results are consistent with the experimentally observed higher tendency of nucleotide-free actin to polymerize [3].

The most significant piece of the puzzle still missing in actin research today is the structure of F-actin at an atomistic level. Detailed knowledge of the structural differences between normal and malignantly transformed actin filaments would allow the design of inhibitors with a high selectivity for actin in tumor cells and avoid side effects related to targeting cytoskeletal components of both normal and cancer cells, such as observed for the microtubule-targeting chemotherapy agent Taxol.

Bibliography

[1] Huang D, Caflisch A. The free energy landscape of small molecule unbinding. PLoS Comput Biol 2011;7:e1002002.

[2] Oda T, Iwasa M, Aihara T, Maéda Y, Narita A. The nature of the globular- to fibrous-actin transition. Nature 2009;457:441–445.

[3] De La Cruz EM, Mandinova A, Steinmetz MO, Stoffler D, Aebi U, Pollard TD. Polymerization and structure of nucleotide-free actin filaments. J Mol Biol 2000;295:517–526.

Acknowledgments

First of all, I would like to thank Prof. Dr. Amedeo Caflisch for the possibility to work in his group, his continuous support and many inspiring discussions. Many thanks to Prof. Dr. Cristina Nevado for a fruitful collaboration, interesting scientific discussions and for being part of my PhD committee. I appreciate the assessment of this work by Dr. Michele Seeber.

I am very grateful to Andrea Unzue for working with me on the actin inhibitor design projects and carrying out the chemical syntheses and binding assays of “our” molecules.

I would like to thank all the members of the Caflisch group for sharing interesting discussions and making the group a pleasant and stimulating place to work at. My deepest gratitude to Emilie Frugier for her friendship and continuous support both on a scientific and personal level. I am very grateful to Tim Knehans for his support and many discussions. Many thanks also to Sandra Steiner, Andrea Magno, François Marchand, Riccardo Scalco, Sabyashachi Mishra, and Danzhi Huang. I would also like to thank the former members of the group Philipp Schütz, Marino Convertino, Pietro Alfarano, and Ran Friedman for scientific and/or technical support. Moreover, I appreciate the efficient technical support by Dr. Stefan Klauser and Steve Rast. Many thanks to Dr. Sergio Gloor for creating a pleasant environment for teaching biochemistry.

Special thanks go to my family, who supported me during all my life, and my boyfriend who never got tired of listening to stories about my work.

I dedicate this work to my beloved father and grandmother who both passed away suddenly and unexpectedly during my time as a PhD student.

Curriculum Vitae

Sandra RENNEBAUM

born 18th October 1984 in Menden (Sauerland), Germany
German citizen

EDUCATION

05/2009–now	Ph.D. thesis at the Department of Biochemistry, University of Zürich, Switzerland.
07/2007–02/2008	Exchange semester at the NTNU Trondheim, Norway.
2005–2009	Studies of Chemistry at the RWTH Aachen University, Germany. Final diploma examination passed with highest honors.
2004–2005	Studies of Chemistry at the TU Dortmund University, Germany.
1995–2004	Heilig-Geist-Gymnasium Menden, final exam passed with highest honors.

Publication during diploma studies:

Hydrophilic carotenoids: facile syntheses of carotenoid oxime hydrochlorides as long-chain, highly unsaturated cationic (bola)amphiphiles.
J. Willibald, S. Rennebaum, S. Breukers, S. H. Hafez, A. Patel, C. L. Øpstad, R. Schmid, S. N. Naess, H.-R. Sliwka, V. Partali
[*Chemistry and Physics of Lipids*, **2009**, *161*, 32-37]